Ollscoil na hÉireann, Corcaigh National University of Ireland, Cork

A GENETIC CODE EXPANSION:

INVESTIGATION OF UGA STOP CODON RECODING IN SELENOPROTEINS

Volume 1 of 1

A thesis submitted to the National University of Ireland, Cork in fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Janinah Marie P. Baclaocos, BSc

August 2019

School of Biochemistry and Cell Biology

Supervisors: Prof. John F. Atkins and Dr. John Mackrill

Head of Department: Prof. Rosemary O'Connor

Table of Contents

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

_____.

List of Figures and Tables

Acknowledgements

Undertaking this Ph.D has truly been a life-changing experience; a ‘marathon’, that would not have been possible to complete without the help and support of many individuals.

I would like to express my sincere gratitude, firstly, to my mentors: Prof. John F. Atkins for his encouragement, mentorship and the many amazing opportunities that was provided for me; to Dr. Gary Loughran for his technical advice, humour and insightful discussions; to Dr. John Mackrill for his patience, guidance and positivity which really made a difference in times of difficulties.

I would also like to thank our many collaborators in the selenium field whose help was crucial to complete this project, especially Dr. Marco Mariotti for providing insightful directions on the project and thorough feedback on the manuscript; Dr. Paul Copeland and lab members for accommodating my short but productive stay in Rutgers and Dr. Didac Santesmasses for always patiently helping me with my data.

Thanks also to my fellow lab members and colleagues, past and present, Martina, Darren, Pramod, Arthur, and members of the LAPTI lab, for sharing knowledge, experience, protocols and techniques.

My sincere gratitude also goes to Sinead, Julia, Anjali, Ioanna and Steph with whom I shared the many ups and downs of this journey. I will always treasure the laughter, friendship and solid science shared over lunch and coffee through the years. My time in Cork has been the most memorable because of our adventures!

I am grateful to my friends outside academia and my family for always supporting and believing in me. Finally, to Regi, who has been extremely patient with the demands of this work but was still always encouraging and there for me when it really mattered, thank you.

List of Abbreviations

Abbreviations of amino-acids

Alanine	**Ala**	**A**
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Selenocysteine	Sec	U
Pyrrolysine	Pyl	O

Abbreviations of selenoproteins

Glutathione peroxidases	**GPX**
Methionine R-Sulfoxide Reductase	MSRB
Methionine S-Sulfoxide Reductase	MSRA
Radical-S-adenosyl methionine	RSAM
Selenophosphate Synthetase	SPS
Selenoprotein J	SELENOJ
Selenoprotein K	SELENOK
Selenoprotein L	SELENOL
Selenoprotein M	SELENOM
Selenoprotein N	SELENON
Selenoprotein O	SELENOO
Selenoprotein P	SELENOP
Selenoprotein S	SELENOS
Selenoprotein U	SELENOU
Selenoprotein V	SELENOB
Selenoprotein W	SELENOW
Selenoprotein W	SELENOW
Thioredoxin Reductase	TXNRD
Thyroid Hormone Deiodinase	DIO

Nomenclature: All higher cases for proteins e.g. SELENOP/SPS; first letter higher case for mRNA e.g. SelenoP/Sps ; and italics for genes e,g. *selenoP/sps* (Gladyshev et al.,2016)

General Abbreviations

3X-F	**3X-Flag tag**
5.8, 5, 18, 28, 30, 60, 80S nomenclature	Svedberg units describing molecular sedimentation
A.mellifera	A. mellifera/honey bee
AAR motif	R for purine nucleotide A/G
APS	ammonium persulfate
ASW	artificial sea-water
ATP	Adenosine tri-phosphate
AUG	annotated start codon/ methionine
Blast	basic local alignment search tool
bp	Basepair
C.gigas	Crassostrea gigas/Pacific oyster
cDNA	complementary DNA
CDS	coding sequence
CT-SBP2	C-terminal half of SBP2
E.coli	Escherichia coli
E.complanata	*Eliptio complanata*/freshwater mussel
EF1A	eukaryotic translation elongation factor 1-alpha
EFSec	Sec-specific eukaryotic factor
eIF4a3	eukaryotic initiation factor 4a3
eIF4E	eukaryotic initiation factor 4E
EJC	exon-junction complex
GFP	green fluorescent protein
GST	glutathione S-transferase
GTP	guanosine tri-phosphate
$H_2SePO_3^-$	Monoselenophosphate
HEK293	Human embryonic kidney cell clone derived from 293rd experiment
HSe	Hydrogen selenide
ICP-MS	Inductively coupled plasma Mass Spectrometry
IRES	Internal ribosome entry site
ISL	Initiation Stem Loop
kDa	kilodalton
K-turn	kink-turn motif in RNA structures
L30	ribosomal protein L-30
L7Ae	kink-turn RNA secondary structure protein binding motif
MgCl2	Magnesium Chloride
mRNA	messenger RNA
MW	molecular weight
NaOH	sodium hydroxide
NMD	nonsense mediated decay
NP-40	Nonidet P-40 non-ionic non denaturing detergent
nt	nucleotide
ORF	open-reading frame
PBS/PBS-T	phosphate buffer saline /phosphate buffer saline-Tween 20
polyA	poly-adenylated tail of mRNA

ppm	parts per million
psTK	phosphotyrosine-kinase
RBD	RNA binding site of SBP2
RF	release factors
ribosomal A-site	binding site of amino-acyl tRNA entry
ribosomal E-site	binding site of de-acylated tRNA exit
ribosomal P site	binding site of peptidyl tRNA
Rnl2	RNA ligase 2
RNPs	ribonucleoprotein complexes
RPF	ribosome protected fragments
RPKM	reads per transcript length (kb) per million mapped reads
RRL	rabbit reticulocyte lysate
rRNA	ribosomal RNA
RT	reverse transcription/reverse transcribed product
SBP2	SECIS binding protein-2
SBP2L	SECIS-binding protein 2-like protein homolog of SBP2
SDS-PAGE	Sodium-Dodecyl Sulphate Polyacrylamide Gel electrophoresis
SECIS	Selenocysteine Insertion Sequence
SecS	Selenocysteine lyase
Sec-tRNA$^{[Ser]Sec}$	charged tRNA for Sec insertion
Sec-UGA	in-frame UGA codon redefined to Sec
Selenium	Se
SEPSECS	selenocysteine synthase
SeRS	seryl-tRNA synthetase
SID	Sec-incorporation domain of SBP2
SRE	Sec-recoding elements
T4 PNK	Polynucleotide Kinase derived from bacteriophage T4
TE	translation efficiency
tRNA	transfer RNA
UGA, UAA, UAG	translation termination codons
UGC	Cysteine codon
Um34	Sec-tRNA$^{[Ser]Sec}$ modification at the wobble position of anti-codon
UTR	untranslated region
WGL	wheat germ lysate

Abstract

After the genetic code was deciphered in the 1960s, Francis Crick formulated the 'frozen accident' hypothesis (Crick, 1968) to describe the origins of the genetic code as universal and resistant to change or evolution. Co-incidentally, evidence of the dynamic nature of genetic decoding emerged through a series of experimental observations which presented various cases of exceptions from what were known as the standard rules of decoding. There is now prevalent understanding and evidence that the genetic code is constantly evolving, and it can be altered by various organisms with possible implications for entire genomes or specific mRNAs.

The incorporation of the 21st amino-acid selenocysteine in selenoproteins in response to the UGA translation 'terminator' codon is an example of a gene-specific expansion of the code. This thesis will deal primarily with two unique cases of UGA recoding. The first case is the synthesis of selenophosphate synthetase 1 (SPS1) (Chapter 2) whereby an unknown amino acid is inserted in response to a UGA codon in the hymenopteran honeybee, *Apis mellifera,* which lacks the machinery for Sec incorporation. The various attempts to characterize the amino acid inserted at this position by novel methods are described.

In Chapter 3, the first extensive evolutionary analysis of the selenium transporting protein, selenoprotein P (SELENOP) in invertebrates is described with focused characterization in the mollusc, Pacific oyster, *Crassostrea gigas*. This unique case presented an unprecedentedly high Sec content (46 Sec) in the C-terminal domain of its SELENOP highlighting an extreme case of deviation from the standard genetic code read-out. It was shown that a supplemented heterologous system, was able to facilitate translation of oyster SelenoP mRNA up to the third or fourth Sec codon position of the distal region but was inadequate to produce the full-length protein. Further, the Sec-dedicated protein factor, the oyster SECIS binding protein 2 (SBP2) was characterized and its potential tested for processive Sec-incorporation. Specific mRNA structures in the 3'UTR, termed Selenocysteine Insertion Sequence (SECIS), are essential for the recoding of UGA to specify selenocysteine instead of termination. While previously known multi-Sec codon SelenoP genes have two functionally distinct SECISes, the two in *C. gigas* showed no distinction *in-vitro*.

In Chapter 4, *in-vivo* selenium regulation of selenoproteins in *C.gigas* was investigated by ribosome profiling. Total selenium levels in oyster tissues were found to increase up to 50-fold with supplementation, also resulting to an increase in mRNA abundance and translation. The translation of the full-length Pacific oyster SelenoP demonstrates an inefficient selenocysteine specification at UGA 1 (> 6%) and very high efficiency at the distal UGAs (UGAs 2 to 46). Additional genetic elements relevant to SelenoP translation include a leader ORF, and the RNA structure, termed Initiation Stem Loop (ISL) which were found to potentially modulate ribosome progression in a selenium-dependent manner. It was further validated that selenocysteines were metabolically incorporated in response to UGAs during the synthesis of oyster SELENOP as indicated by ^{75}Selenium labelling experiments. These findings highlight the increasing understanding of the plasticity of the genetic code, as well as the ecological importance of selenium and its diverse utilization across species.

Chapter 1:

General Introduction

1.1 : The genetic code variation

The general nature of the genetic code readout was elucidated by Crick et al., in 1961 and was described as starting from a 'fixed point' followed by sequential reading in non-overlapping units of three nucleotides (Crick et al., 1961). Based on experiments performed in *E.coli* , the correspondence of particular codons with the encoded amino acids was established (Nirenberg et al., 1963). All 64 triplets of nucleotides or codons were assigned to the 20 naturally occurring amino acids or a specialized function during translation (Nirenberg et al., 1966). The codon, AUG, is recognized to have a dual function serving to indicate translation initiation or methionine insertion within protein coding sequences. Three codons, UAG, UAA and UGA were assigned as signals to terminate protein synthesis.

The characterization of the genetic code at the time, led to the assumption that the code is not only universal but also fixed or 'frozen' and resistant to evolution (Crick, 1968). However, several major variations of the genetic code were reported since the mid-1960's leading to a 'melting' perception of the genetic code's universality (Riyasaty and Atkins, 1968, Weiner and Weber, 1971, Barrell et al., 1979, Horowitz and Gorovsky, 1985, Jacks and Varmus, 1985, Preer et al., 1985, Chambers et al., 1986, Craigen and Caskey, 1986, Huang et al., 1988, Srinivasan et al., 2002, Prat et al., 2012, Lang et al., 2014). Such variations presented potential latitude in both the mode of its readout and especially in the function or amino acid identity specified by one or more codons.

Currently, there are numerous reports of organelles and certain organisms that use a different genetic language. An examination of The NCBI Genetic Code database provides a list of 32 distinct genetic code tables. Most tables in the list are derived from the mitochondria of different species demonstrating variant codes. It has been shown that the mitochondria demonstrate frequent codon reassignment where the meaning of the codons is globally altered (Knight et al., 2001). There are also non-organelle cases involving reassignment of stop codons. For instance, in mycoplasma, UGA is decoded as tryptophan (Yamao et al., 1985) while in *Euplotes*, UGA codes for cysteine instead of termination (Meyer et al., 1991). Until recently, the only known cases of sense codon reassignment is of several species of the *Candida* genus that use

CUG to code for serine instead of leucine (Pesole et al., 1995, Kawaguchi et al., 1989) and in *Pachysolen tannophilus*, a yeast related to the Pichiaceae where CUG codes for alanine instead of leucine (Muhlhausen et al., 2016, Riley et al., 2016).

The characterization of non-universal amino acids: Selenocysteine (Sec) (described in detail in the subsequent sections) and Pyrrolysine (Pyl) presents another example of a genetic code expansion. Sec is incorporated in response to the UGA stop codon in most bacteria and in all archaea and eukaryotes, mediated by cis and trans-acting signals in selenoprotein mRNAs (Bock, 2000). The exception to this was identified in certain bacterial species where the cysteine codon UGU in *Aeromonas salmonicida* bacteria or the UAG stop codon in *Geodermatoophilus* and *Blastococcus* are utilized to insert Sec (Mukai et al., 2016). Pyl is inserted in response to the UAG in archaea and methanogenic bacteria (Srinivasan et al., 2002). In contrast to Sec-incorporation, Pyl interacts with standard elongation factor and does not require specialized RNA structure (Namy et al., 2007), although there are reports that RNA structure affects Pyl insertion efficiency (Longstaff et al., 2007).

The expanded genetic code is shown in Fig 1.1, where Sec is the 21st amino-acid and Pyl is the 22nd amino acid (Srinivasan et al., 2002, Hao et al., 2004). The possibility of a 23rd amino acid has been suggested but deemed to be unlikely (Lobanov et al., 2006). In addition, if a 23rd amino acid exists, such occurrence would be very rare compared to that of Sec and may only be limited to certain organisms. Moreover, another variation of the genetic code was presented where UGA was shown to specify both Cys and Sec incorporation in some Euplotes within a single mRNA. The precise mechanism for discrimination is unclear but evidence showed that it might be dependent on the location of the UGA codon within the mRNA and its distance from stimulatory structures (Turanov et al., 2009, Turanov et al., 2013).

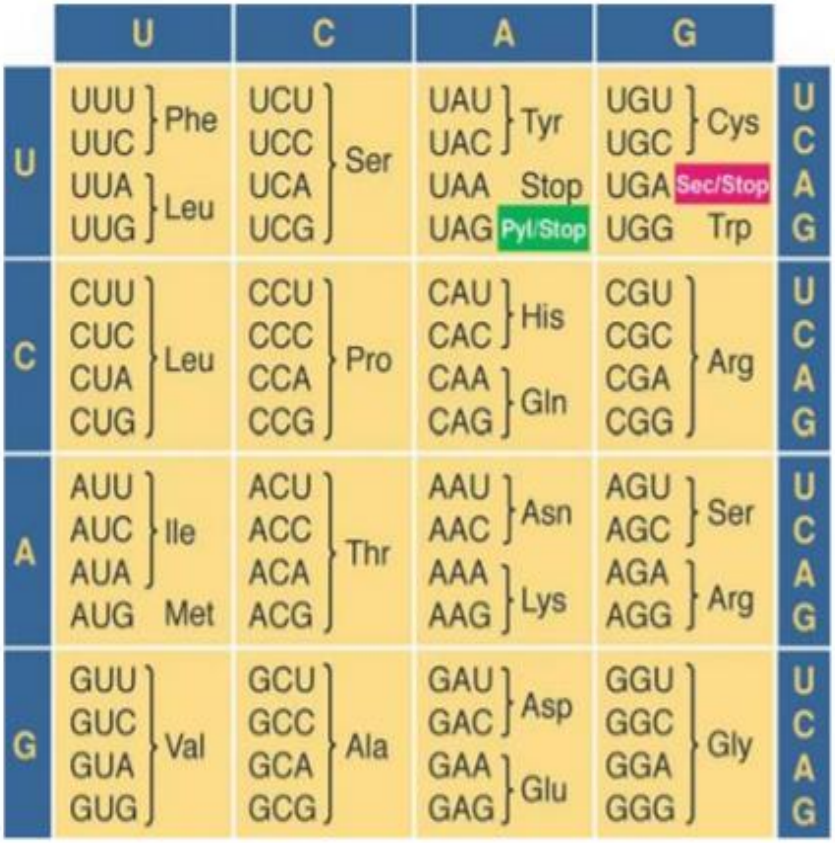


Figure 1.1: The Expanded Genetic Code.

The 64 combination of codons (triplet nucleotides) are presented along with their corresponding amino acid. UGA is included as coding for Sec (found in bacteria, archaea and eukaryotes) or Stop and UAG as Pyl (found in bacteria and archaea) or Stop (Atkins and Gesteland, 2010).

1.2: Recoding mechanisms in the context of stop codon meaning

The process of standard translation termination ensues when the ribosome encounters a stop signal on the mRNA resulting to the release of the completed protein, ribosome dissociation and recycling. Stop codons UAA, UAG or UGA are recognized by release factor proteins (RF1 and RF2 in bacteria with differing specificities (Scolnick et al., 1968, Capecchi, 1967) and the omnipotent eRF1 in eukaryotes (Frolova et al., 1994)). The flexibility of the genetic code is often demonstrated by the involvement of stop-codons either reassigned globally in the genome or 'recoded' context specifically by signals contained within an mRNA.

For simplicity, we group cases of genetic code variation discussed into 'codon reassignments' and 'recoding' (Atkins and Baranov, 2010, Baranov et al., 2015). Codon reassignment present changes in codon meaning on the scale of an entire genetic code. The new meaning of the codon is altered irrespective of mRNA context and applies to the entire mRNA repertoire. Such cases have been described briefly in Section 1.1 and are present in bacterial and specific mitochondrial genomes (Knight et al., 2001). They usually involve changes in translational apparatus such as alteration or loss of a release factor (Duarte et al., 2012). The insertion of the 22nd amino-acid, Pyl, has features of both recoding and reassignment partially depending on whether bacteria or archaea is involved. For instance, in Pyl incorporation, a dispensable RNA secondary structure (Namy et al., 2007) has been identified although it has been reported that the efficiency of Pyl incorporation is affected by certain contexts of such RNA structure (Longstaff et al., 2007). In addition, the extreme scarcity of the UAG codons in Pyl-utilizing archaea and alterations in the mRNA recognition domain of their release factors suggest that UAG is mostly utilized during Pyl incorporation as opposed to a termination signal (Zhang et al., 2005a).

In contrast, 'recoding' encompasses the process of context- or condition-specific alteration of genetic decoding (Baranov et al., 2015). They are usually dynamic as they are in competition with standard decoding, with some of the product reflecting standard meaning while others resulting to a novel protein with a different function (Atkins and Baranov, 2010). Recoding signals within an mRNA or nascent peptides are usually required and, in some cases, additional components of the translation apparatus are necessary. Such events are also dependent on cellular

conditions (Gesteland et al., 1992, Atkins and Gesteland, 2010, Baranov et al., 2015). Many characterized recoding events involve stop codons UAG, UAA and UGA highlighting their contribution to the genetic code's versatility and plasticity. We focus on examples of recoding where stop-codon meaning is altered during translation and describe the current mechanisms proposed underlying these events. Such examples include programmed +1 ribosomal frameshifting, translational bypassing in T4 bacteriophage *gene 60*, stop codon readthrough and Sec-redefinition. The mechanism of Sec-redefinition will be discussed in detail separately in the subsequent sections.

Programmed ribosomal frameshifting

In normal translation, ribosomes are expected to decode proteins as continuous sequences of nucleotide triplets to a corresponding amino-acid. Programmed ribosomal frameshifting describes the process whereby the ribosome 'shifts' usually at +1 or -1 nucleotide position(s). This results to a change in mRNA open reading frame (ORF) translated by the ribosome and the production of alternative proteoforms, often with differing functions (latest comprehensive review in (Atkins et al., 2016)

In some cases, frameshifting can be defined by the requirement of a 'frameshift site'. This is determined by a specific mRNA sequence where a frameshift could occur and a 'stimulatory element' which describes the sequence within the same mRNA that improves the efficiency of the frameshift. Stimulatory elements of diverse types exist in different organisms. In bacteria, mRNA complementary to ribosomal RNA (rRNA) can stimulate frameshift at +1 or -1 direction (Weiss et al., 1988, Larsen et al., 1994, Prere et al., 2011). Other stimulators include nascent peptides (Yordanova et al., 2015, Gurvich et al., 2011), RNA secondary structures like stem-loops (Kim et al., 2014, Yu et al., 2011), simple (Brierley et al., 1989) and complex pseudoknots (Baranov et al., 2005, Plant et al., 2005), triple helices (Chou and Chang, 2010, Chen et al., 2009), kissing loops (Herold and Siddell, 1993), G-quadruplexes (Endoh and Sugimoto, 2013, Yu et al., 2014) and long-range interactions (Tajima et al., 2011, Barry and Miller, 2002). Interaction of mRNA with various cellular components may alter the stimulatory properties of such structures and sequences (Howard et al., 2004, Olsthoorn et al., 2004, Li et al., 2014).

Frameshift cases where stop codons are utilized include +1 frameshift to produce the bacterial release factor 2 (RF2). This recoding event occurs mostly at

nucleotide context C.CU_**U.GA** (Baranov et al., 2002b) (underscore denotes codon boundaries at the initial reading frame and dots denote boundaries at the shifted frame). In addition, the pairing between a specific sequence 5′ of the shift site with the rRNA of translating ribosomes is required. Also, the ornithine decarboxylase (ODC) gene that generates antizyme by +1 frameshifting requires the UGA stop codon and an RNA pseudoknot at the 3'UTR (Ivanov et al., 2000). These events are usually highly responsive to cellular conditions and provides translation with a powerful regulatory mechanism. In *Euplotes* spp., pervasive frameshift occurs in 10% of its genes. Such events are strongly associated with the sequence A**.** AA_**U. AA**_N or its minor variant A .AA_**U.AG**_N (Klobutcher and Farabaugh, 2002, Lobanov et al., 2017). The *Euplotes crassus* genetic code have also reassigned UGA codon as cysteine (Turanov et al., 2009). The current hypothesis for the high frequency of frameshifting is the altered mRNA recognition properties that weakens UAG and UAA codon recognition by its own release factors. Inefficient termination at this codon may likely trigger ribosomal frame shifting at AAA codons.

Translational bypassing

The translation of the bacteriophage T4 gene *60*, also known as ribosomal hopping, results to a product where the two adjacent codons are separated by a 50-nucleotide (+50) non-coding gap. Translation is disrupted at a specific glycine codon and subsequently resumes 50 nucleotides downstream at a codon 3' adjacent to another glycine codon. Studies performed to elucidate the mechanism of translational bypassing revealed a complex consecution of stimulatory elements within the mRNA. Among these, are a required stop codon at the beginning of the non-coding gap that could be dynamically folded into an RNA secondary structure and a specific sequence of the nascent peptide (Wills et al., 2008, Herr et al., 2000a, Herr et al., 2000b, Weiss et al., 1990).

Stop codon readthrough

Stop codon readthrough extends distal region of the same mRNA from a stop-codon to synthesize an extended second protein product. In viruses, this mechanism increases the functional versatility of a compact genome and controls the ratio of two protein isoforms (Namy and Rousset, 2010). Reports of stop-codon readthrough were previously rare in cellular chromosomal genes (Dreher and Miller, 2006, Steneberg

and Samakovlis, 2001, Namy et al., 2003, Robinson and Cooley, 1997, Klagges et al., 1996). However, emerging studies have shown that readthrough events are abundant in some species during translation. Analysis of protein coding genes in *Drosophila* species predicted prevalent UGA stop-codon readthrough in almost 300 genes (Jungreis et al., 2011) verified by ribosome profiling experiments (Dunn et al., 2013). The protein products from these genes were implicated in fruit-fly fitness and are likely to be expressed preferentially at certain stages of development. In addition to prevalent cases in *Drosophila*, cases of stop-codon readthrough have also been documented in mammalian genes (Loughran et al., 2014, Eswarappa et al., 2014, Schueren et al., 2014, Stiebler et al., 2014).

Short nucleotides sequence constraints (as short as 6 nucleotides) located downstream of a stop codon can stimulate significant readthrough efficiency higher than background levels (Skuzeski et al., 1991) . In some genes, higher levels of stop-codon readthrough are facilitated by additional elements such as RNA secondary structures (Firth et al., 2011)

Selenocysteine (Sec) incorporation

Translation of selenoprotein mRNAs requires the insertion of the 21st amino-acid Sec in response to an in-frame UGA codon. This involves Sec biosynthesis on its own tRNA and a regulated interaction of the selenoprotein mRNA with a specialized decoding apparatus during translation. Sec insertion is highly dependent on the trace element selenium. The succeeding sections will focus on the characterization of Sec incorporation during selenoprotein synthesis.

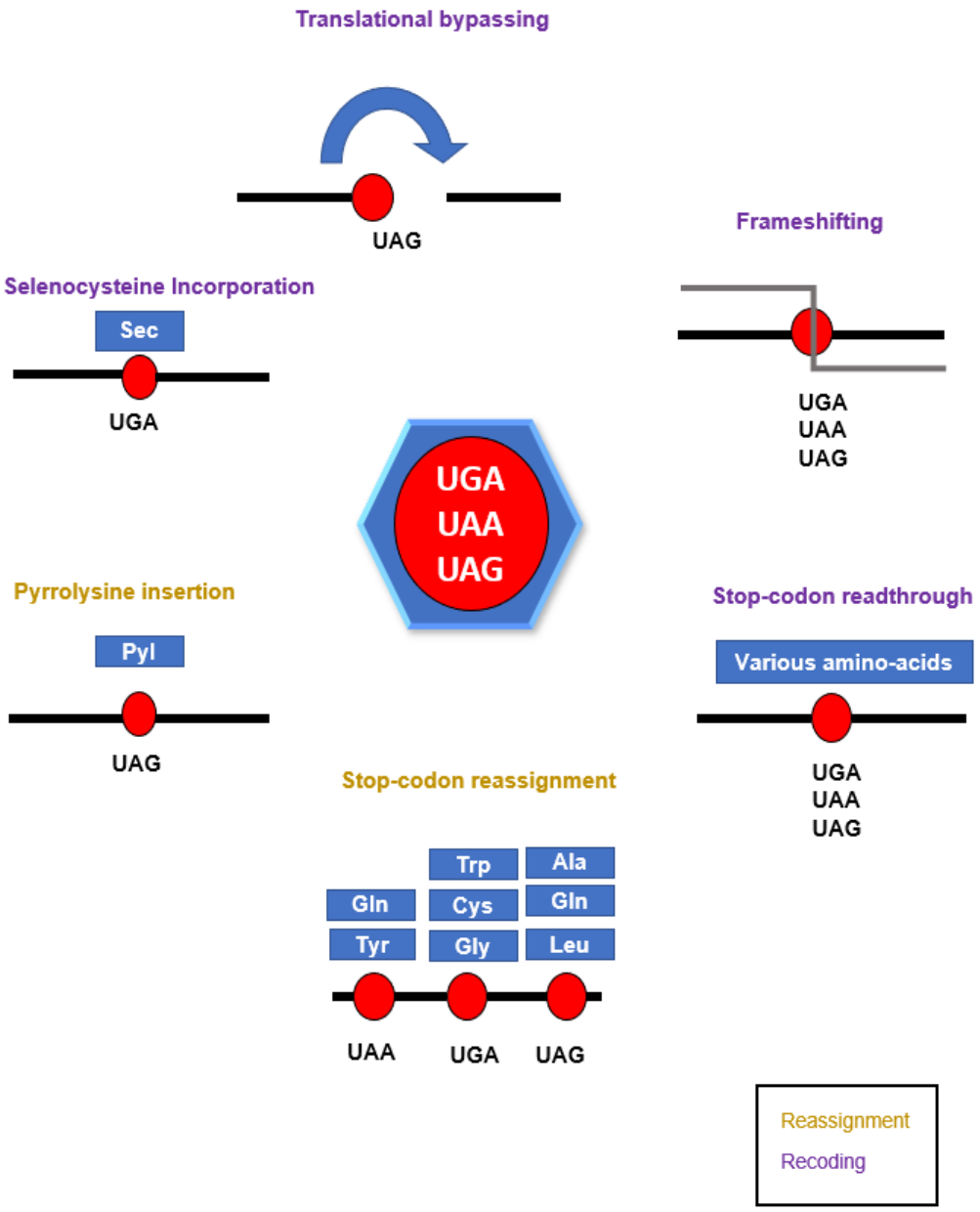


Figure 1.2: Stop codon alteration during reassignment and recoding events

Stop codons (red circles) are reassigned as various amino acids (blue) during codon re-assignment (orange). Various case-specific recoding events (purple) are also dependent on stop codons as well as stimulatory signals or conditions e.g. stop-codon readthrough, +1 frameshifting, bypassing and selenocysteine incorporation. Adapted from Baranov et al., 2015. (Baranov et al., 2015)

1.3: The nutritional selenium requirement

Selenium (Se) is now known to be an essential micronutrient in many living organisms including humans. The element was discovered by Berzelius in 1818 as a by-product of sulphur production and shares various properties with the element tellurium. Se can occur in the environment as inorganic (elemental Se, selenite and selenate), and organic chemical species, including methylated Se compounds and seleno-amino acids (See Section 1.4). Many of the diverse functions of Se are exerted through the biological incorporation of Sec in selenoproteins. Despite the recognized essentiality of Se, a narrow window exists between basic human Se requirement allowing optimal function, and levels of toxicity or deficiency (Mertz, 1972)

Deleterious effects were originally associated with Se as reports emerged of a fatal disease amongst horses grazing in certain areas of the USA and China in the 1940s (Moxon and Rhian, 1943). Isolated cases of Se-poisoning in forage-eating animals such as cattle, sheep and horses were recorded indicating the importance of monitoring Se levels in soils. Se toxicity or selenosis is characterized by blood levels of 100 μg/dL resulting from extended consumption of high amounts of Se (Koller and Exon, 1986). Affected individuals exhibited symptoms of hair- and nail-loss, skin lesions and a dysfunctional nervous system (Yang et al., 1983). On the other hand, two diseases, Keshan disease which results to chronic myocarditis (Jun et al., 2011) and Kashin-Beck disease, a multi-factorial bone and joint disease (Zhou et al., 2014), have been associated with severe Se deficiency. Both diseases respond to prophylactic oral Se administration (Ge et al., 1983, Moreno-Reyes et al., 2003). Studies have now recognized Se deficiency as a contributing factor in pathophysiological conditions including heart disease (Handy and Loscalzo, 2012), neuromuscular disorders (Moghadaszadeh et al., 2001), cancer (F. Jr. Combs and Lu, 2012), male infertility (Turanov et al., 2012) and inflammation (Kaushal et al., 2012). In addition, Se also plays a role in mammalian development (Köhrle, 2000) and immune functions (Rayman, 2012).

Clinical Sestatus is measured by assaying selenoproteins in serum or whole blood (Sunde, 2010, Xia et al., 2010). Plasma or serum levels of glutathione peroxidase 3 (GPX3) and selenoprotein P (SELENOP), are highly informative biomarkers of Se status in humans (Xia et al., 2005, Kipp et al., 2015), with the latter

accounting for 60% of Se in human plasma, due to its role in Se transport and metabolism. The geographical variation of Se in soil and subsequently in plants and animals consumed by humans, reflects the variation across inhabitants of such locations. Lower Se intake, and thus a lower Se status has been reported for Germany and many other European countries, whereas the USA and Mexico exhibit higher Se status in their populations (Combs, 2001). Such variation across countries are reflected by the Recommended Daily Amounts (RDA) for Se consumption e.g. 75μg/day in Germany and 55 μg/day in the US. Se intake also varies across populations of a country, with China as a prime example of both extremely low (16±4 μg Se/L in plasma or serum) or extremely high intakes (1438.2 ± 76.3 μg Se/L) (Xu and Jiang, 1985). Information on Se status across geographical locations is important to better assess Se effects and prevent cases of severe toxicity or deficiency.

Since Se deficiency is a recognized problem world-wide with negative impacts on individual human health and lifespan, the requirement to increase Se consumption becomes apparent. Biofortification, the process of increasing the nutritional quality of food, have gained popularity to increase Se consumption by humans and animals (Ramos et al., 2010). Across the globe, innovative technological processes to enrich products such as eggs, pork, poultry and milk have been successfully introduced (Navarro-Alarcon and Cabrera-Vique, 2008, Finley, 2006).

1.4: Dietary forms of selenium and their metabolism

The main chemical forms of bioavailable selenium are selenomethionine (SeMet) and selenocysteine (Sec) and the inorganic forms selenite and selenate. These forms account for almost all the Se in the diet and are readily absorbed without regulation.

Inorganic Selenium

Two commonly forms of inorganic Se used in supplementation are selenite and selenate. Both are efficiently absorbed but selenate needs to be reduced to selenite once absorbed before it can be further metabolized (Dael et al., 2007). Subsequently, selenite is converted to selenide by thioredoxin reductases (TrxR) (Kumar et al., 1992) or glutathione interactions (Ganther, 1971) in the intestinal mucosal cells. Selenide is the form that is utilized for selenoprotein synthesis. Proteins that bind selenium administered as selenite were identified in mouse liver cytosol (Bansal et al., 1989). In *Arabidopsis thaliana* (mouse-ear cress), a Se-binding protein homolog was identified after selenite supplementation and was observed to form selenotrisulfides, in between two cysteine residues (Schild et al., 2014). Selenotrisulfides were initially characterized as a non-enzymatic type of Se binding which occurs *in-vitro* when selenite and thiols are present together (Ganther, 1968). The relevance of such selenium-binding proteins to selenium metabolism is still to be determined.

Selenomethionine (SeMet)

The principal form of selenium in most human and animal diets is SeMet. It is biosynthesized by plants and incorporated in their proteins at methionine positions. About 90% of Se in plants are present as SeMet while the remainder mostly accounts for Se-containing analogs of other sulfur compounds (Cubadda et al., 2010, Olson et al., 1970).

SeMet is absorbed via intestinal methionine transporters (McConnell and Cho, 1967, Wolffram et al., 1989) and enters the pool of free methionine in the body. From this pool, SeMet can be randomly inserted into proteins at methionine positions. A ratio of 1 SeMet molecule per 1,100 albumin molecules has been reported in the albumin of a healthy human subject in a US study (Burk et al., 2001). Degradation of SeMet containing proteins allows SeMet release to the free methionine pool. SeMet

can also be metabolized, mostly in the liver via the methionine cycle and transsulfuration pathway yielding selenocysteine (Esaki et al., 1982).

The γ-Lyase enzyme has been reported to act on SeMet to produce methylselenol (Okuno et al., 2005), the biological significance of which with subsequent demethylation to selenide, is still undetermined (Suzuki et al., 2007). The only known biological function of SeMet in proteins is to serve as an unregulated reserve source for selenium (Waschulewski and Sunde, 1988). During turn-over of the methionine pool, Se is freed and enters the intracellular pathway for the synthesis of Sec and subsequent incorporation into proteins (see Fig 1.3). Thus, SeMet metabolizing organisms will continue to have Se-supply for selenoprotein synthesis for a limited time while exposed to a low Se environment.

Selenocysteine (Sec)

Free selenocysteine occurs in plants as an intermediate in the reversible transsulfuration pathway that produces SeMet (Sors et al., 2005) and occurs less than SeMet (Schild et al., 2014). Its detection in tissue homogenates has not been reported suggesting that its abundance in tissues is low (Esaki et al., 1981) in comparison to SeMet which is metabolized as if it were methionine. Such concentrations do not achieve efficient incorporation to cysteine tRNA, and once incorporated directly into Cys-containing proteins, increase their predisposition to oxidative damage leading to degradation (Sabbagh and Van Hoewyk, 2012). Some Se-accumulator plants eliminate free Sec by methylation to produce Se-methylselenocysteine which cannot be incorporated into proteins (Neuhierl et al., 1999). Se-methylselenocysteine is bioavailable but can be toxic upon consumption of plants that metabolize this compound. Sec from SeMet can also enter the intracellular Se-cycle upon distinction of the Sec lyase enzyme from its sulfur analog, cysteine, and subsequent conversion to selenide (Fig 1.3)

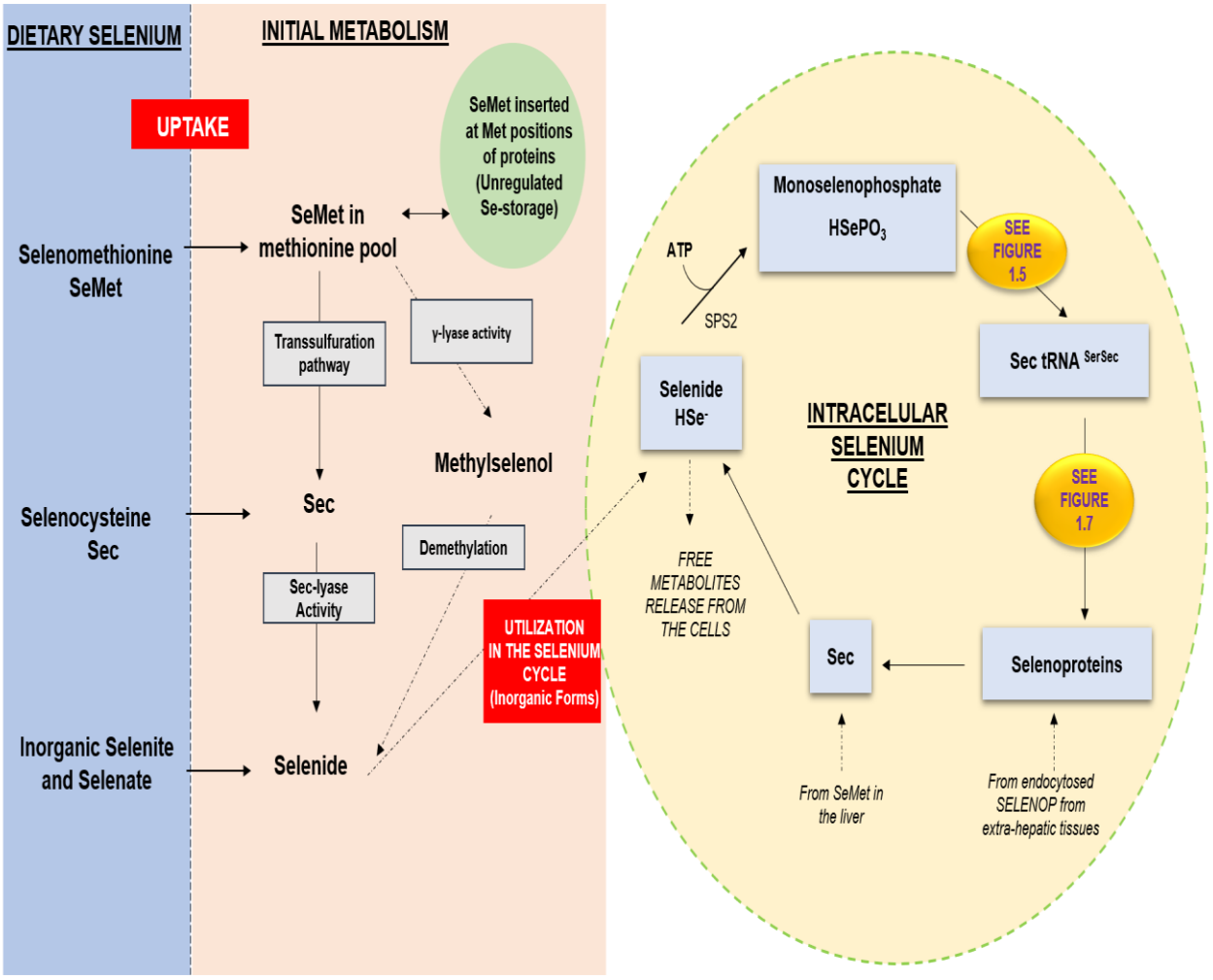


Figure 1.3: Metabolic pathway of selenium forms and the selenium cycle to selenoprotein synthesis.

Selenomethionine enters the methionine pool and can be randomly inserted at methionine positions of proteins. When it enters the transsulfuration pathway, selenocysteine becomes an intermediate and is distinguished by the enzyme selenocysteine lyase. Sec is converted to selenide that can enter the selenium cycle for selenoprotein synthesis. Inorganic selenite/selenade is converted to selenide after absorption and are utilized in the Se-cycle. In the Se-cycle (yellow), the Se intermediates are further metabolized ultimately leading to Sec tRNA biosynthesis (Fig. 1.5) needed to synthesise selenoproteins (Fig 1.6). Adapted from Burk and Hill 2015 (Burk and Hill, 2015).

1.5 The biosynthesis of selenocysteine

The Se-containing amino-acid selenocysteine is distinct to any other amino-acids in eukaryotes as it is biosynthesized on its own tRNA, designated as Sec-tRNA$^{[Ser]Sec}$ (Turanov et al., 2011). In this section, we describe the unique features and modifications of Sec-tRNA $^{[Ser]Sec}$ and the pathway to Sec biosynthesis.

Selenocysteine (Sec)-tRNA $^{[Ser]Sec}$

The tRNA$^{[Ser]Sec}$ gene, designated *Trsp,* usually occurs as a single gene copy in all organisms examined except for zebrafish, which contains two gene copies (Xu et al., 1999). Gene inactivation of *Trsp* (*ΔTrsp)* leads to embryonic lethality in mice (Bosl et al., 1997) suggesting a critical role of selenoproteins in mammals. The presence of this gene in the genome is the most conclusive sign of selenoprotein expression in an organism.

Several unique features of Sec- tRNA$^{[Ser]Sec}$ structure distinguishes it to other classical tRNAs (See Fig 1.4) (Carlson et al., 2016). Such features are found in its 2D-structure comprising of nine base-pairs (bps) in its acceptor arm and four bps in its TψC arm. This results to a 9/4 cloverleaf form compared to a 7/5 form found in other tRNAs. The D-arm of Sec tRNA also contains more bps (five to six versus three to four in other tRNAs) and a longer variable arm (16 bps versus 4 to 5 bps in others). Also, Sec-tRNA does not have a dihydrouracil base usually found in canonical tRNAs. Together, these features make Sec-tRNA the longest sequenced tRNA in higher vertebrates and the most unique adaptor RNA described to date (Böck et al., 1991, Sturchler et al., 1993). The unique structure of Sec-tRNA results in a steric inhibition of interactions with the canonical elongation factors, EFTU in prokaryotes and EF1alpha (eEF1A) in eukaryotes. Thus, dedicated elongation factors for Sec-incorporation are required namely, elongation factors SelB in bacteria (Fagegaltier et al., 2000a) and eEFSec in eukaryotes (Tujebajeva et al., 2000).

Further modifications on tRNA$^{[ser]sec}$ occurs upon maturation in four positions namely 5-methoxycarbonylmethyluridine (mcm^{5}u), N6-isopentenyladenosine (i^{6}A), pseudouridine (ψ) and 1-methyladenosine (m^{1}A) at nucleotide positions 34, 37, 55 and 58 respectively (Diamond et al., 1993). These modifications are significantly fewer than in classical tRNAs that normally contain 15 to 17 amended bases. In mammals,

the Sec-tRNA$^{[ser]sec}$ population consists of two isoforms that differ by a single 2'-*O*-methyl moiety in the uridine position 34, designated Um34 or mcm^{5}Um, located at the wobble position in the anti-codon arm (Hatfield et al., 1991). The non-Um34 isoform is involved in the synthesis of house-keeping selenoproteins while the Um34 isoform governs the synthesis of stress-related selenoproteins (Carlson et al., 2005, Carlson et al., 2007). Modification at Um34 is facilitated by the mammalian tRNA methyl transferase ALKBH8, along with an accessory protein, TRM112 (Songe-Moller et al., 2010). The extent of the Um34 modification is dependent on selenium levels which in turn correlates to the expression of stress-related selenoproteins (Diamond et al., 1993, Chittum et al., 1997, Hatfield et al., 1991).

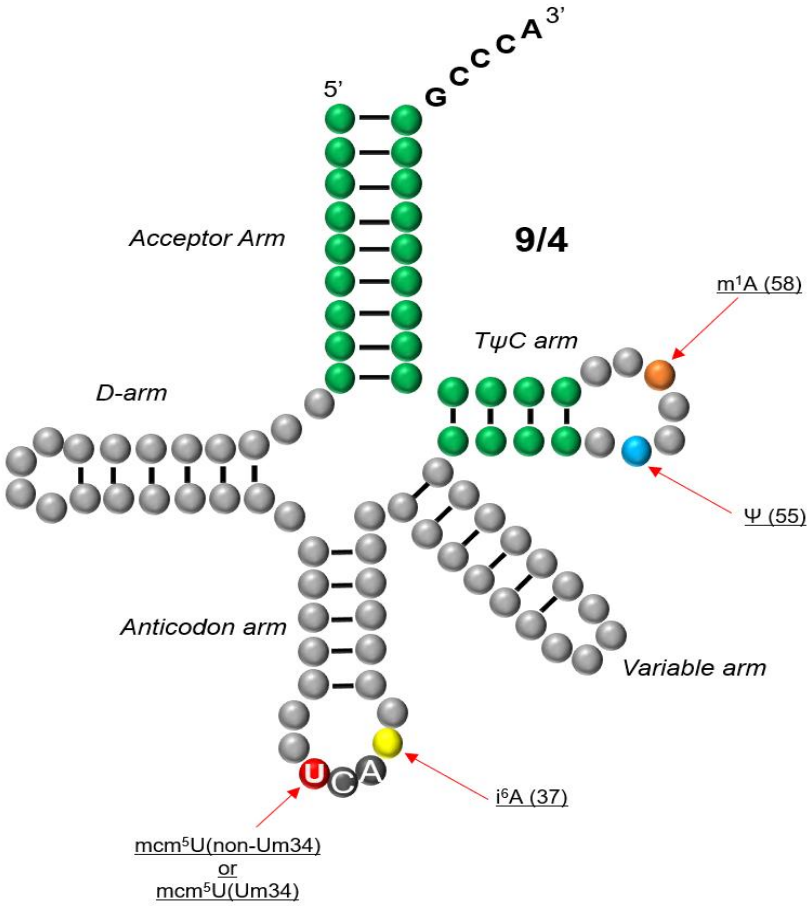


Figure 1.4: Selenocystenine (Sec) tRNA$^{[ser]sec}$ cloverleaf model

The acceptor arm constitutes the 5' and 3' paired terminal bases and the D-arm contains four unpaired and six paired bases of the left portion of the tRNA. The anticodon arm contains five paired and 7 unpaired bases containing the Sec-anticodon. The uracil in the wobble position 34 contains the mcm^{5}U at position 34 modification which can undergo further methylation generating mcm^{5}Um (or Um34) and the i^{6}A modification at position 37. The variable arm contains 6 paired bases and four unpaired bases and the TψC arm contains 4 paired and 7 unpaired bases which holds the m^{1}A and ψ modifications at positions 58 and 55 respectively.

Selenocysteine biosynthesis pathway

Sec biosynthesis on $tRNA^{[ser]sec}$ requires a series of enzymatic reactions which involve four enzymes rather than one dedicated amino acid-tRNA synthetase, as for classical tRNAs (see Fig 1.5). The process begins by aminoacylation of $tRNA^{[ser]sec}$ with serine in the presence of seryl-tRNA synthetase (SeRS) (Lee et al., 1990). The serine moiety on Ser- $tRNA^{[ser]sec}$ is subsequently phosphorylated by phosphoseryl-$tRNA^{[ser]sec}$ kinase (PSTK) to form the intermediate, *O*-phosphoseryl-$tRNA^{[ser]sec}$ (PSer-$tRNA^{[Ser]Sec}$) (Carlson et al., 2004). Se in the form of selenide (HSe^-) and the enzyme, selenophosphate synthetase 2 (SPS2 or SEPHS2) faciltates the synthesis of the active selenium donor, monoselenophosphate (H_2SePO_3). SPS2 is a unique enzyme in the Sec-synthesis pathway as it is a selenoprotein itself (Guimaraes et al., 1996, Kim et al., 1997). In mammals, two gene copies of *sps* were identified, defined as *sps1* and *sps2*. Only SPS2 has been implicated in the Sec-biosynthesis pathway while SPS1 may have a function independent of Sec or selenium (details are discussed in Chapter 2). Finally, the pSer-$tRNA^{[ser]sec}$ is converted to a dehydroalanyl-$tRNA^{[ser]sec}$ intermediate where the donation of monoselenophosphate forms the Sec-$tRNA^{[ser]sec}$, with both steps facilitated by Sec synthase (SEPSECS) (Xu et al., 2007b).

Sulfide can replace selenide in the same reaction which involves SPS2 in the *de novo* biosynthesis of cysteine (Cys) (Xu et al., 2010). This forms thiophosphate which can then form Cys- $tRNA^{[ser]sec}$. This tRNA can in turn insert a Cys amino-acid in response to UGA in selenoprotein mRNAs. Cys to Sec replacement occurs naturally *in-vivo* and is also dependent on selenium availability (Lu et al., 2009). Therefore, Sec and *de-novo* Cys differs only in the presence of a selenium versus a sulfur atom respectively. Despite their high similarity, key differences in their chemistries make Sec unique to Cys. Selenium is highly polarizable which makes Sec a better nucleophile and electrophile in comparison to Cys (Hondal et al., 2013, Steinmann et al., 2010). Selenol of Sec has a lower pK_a (pK_a 5.5) compared to the thiol of Cys (pK_a 8.7) (Huber and Criddle, 1967) which means that Sec is almost completely deprotonated at physiological pH further enhancing its nucleophilicity. Such properties are necessary in the functions of selenoproteins in catalyzing reduction/oxidation reactions which are important in the protection of cell membranes, proteins, and nucleic acids from cumulative oxidative damage (Reviewed (Steinbrenner et al., 2016))

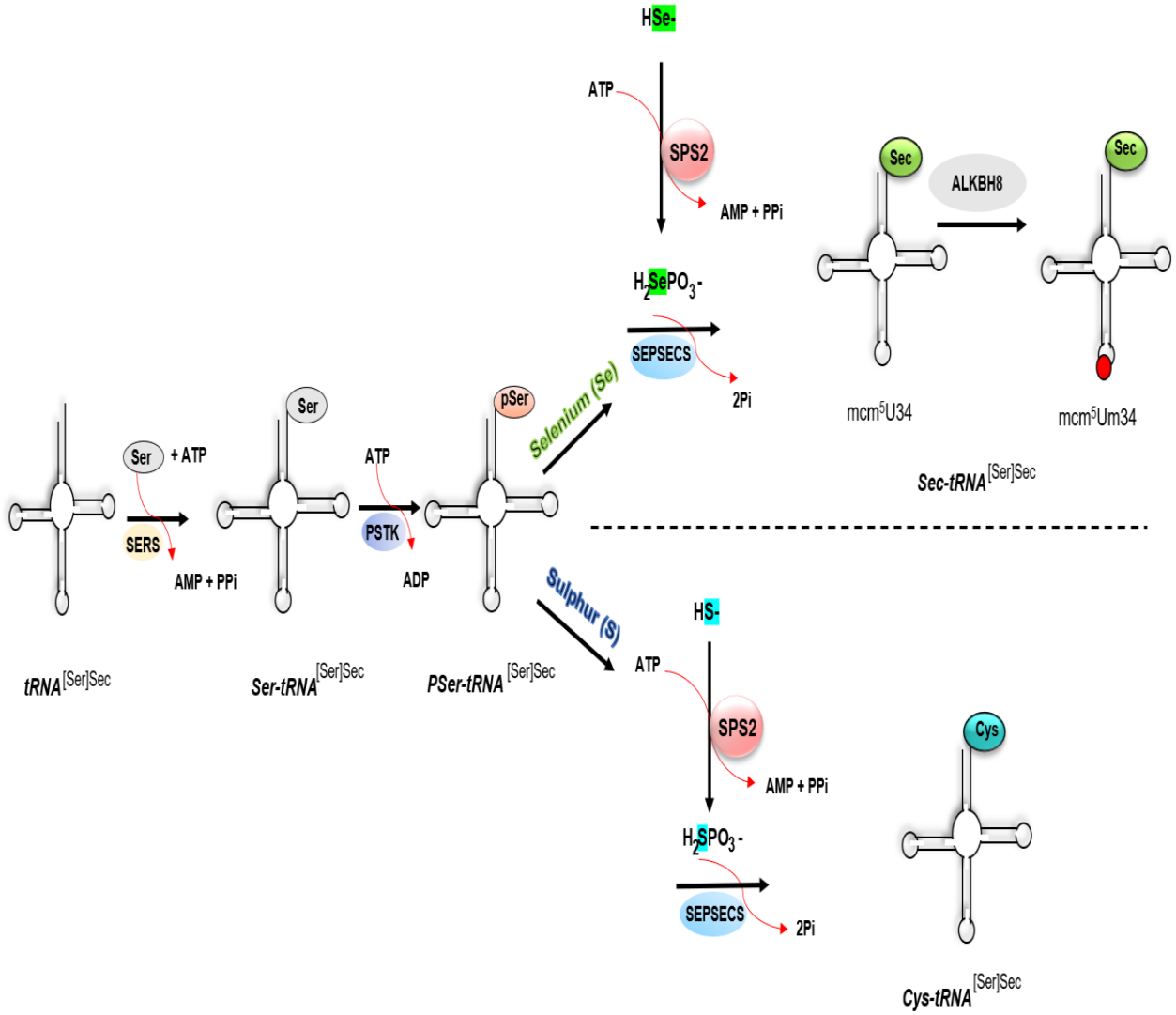


Figure 1.5: Biosynthesis of selenocysteine and de-novo cysteine

Sec is biosynthesized by conversion of its premature $tRNA^{[ser]sec}$ to various intermediates facilitated by various enzymes at each step (details are in the text). Ultimately, the presence of selenium leads to selenophosphate synthetase 2 (SPS2) conversion of selenide (HSe^-) to monoselenophosphate (H_2SePO_3-) which is the active donor of Sec to the mature Sec-tRNA catalyzed by Sec-synthase (SEPSECS). The two mature Sec-tRNA isoforms are shown (mcm^5U34 and mcm^5Um34). Um34 modification is catalyzed by the methyl transferase ALKBH8 (top right corner). The alternative pathway to de-novo cysteine synthesis is also depicted (bottom right corner) where sulfur is used in a similar reaction.

1.6: Molecular Mechanism of Eukaryotic Selenoprotein Synthesis

The molecular machinery for the incorporation of Sec in eukaryotic selenoproteins requires cis-acting factors within selenoprotein mRNAs and trans-acting factors involved in the transport of Sec-tRNA$^{[ser]sec}$ to the ribosome in-response to the UGA codon (Tujebajeva et al., 2000). In this section, we discuss individual components and relevant features of the Sec-insertion machinery. The mechanism of translation that governs most selenoprotein with a single in-frame UGA codon will be described. The exceptional case of SELENOP synthesis, defined by the requirement of highly processive redefinition of multiple UGAs is discussed separately in Section 1.7.

Selenocysteine (SEC) Insertion Sequence (SECIS) Elements

Selenoprotein mRNAs can be distinguished from other mRNAs by the presence of an in-frame UGA codon as well as an RNA secondary structure, termed SECIS element, in its 3'UTR (Guimaraes et al., 1996, Berry et al., 1991). SECIS elements are highly variable across species but contain signature motifs described below that allow for distinction.

In eukaryotes, the SECIS elements are defined by the formation of two helices separated by an internal loop, a quartet of GA nucleotides forming the SECIS core and an apical loop (Walczak et al., 1996). Some SECIS elements exhibit an additional mini-stem (forming an internal bulge) which classifies the SECIS elements into two different types. Type 1 SECIS elements have a larger apical loop and are lacking the mini-stem while type 2 SECIS elements contain the mini-stem (See Fig 1.6) (Fagegaltier et al., 2000b).

The GA quartet is located at the base of Helix II and consist of non-Watson-Crick base pairs, including G-A/A-G pairs, characteristic of kink-turn (K-turn) motifs in RNA secondary structures (Goody et al., 2004, Matsumura et al., 2003). This quartet is highly conserved and interacts with the K-turn motif binding domain or L7Ae domain of the dedicated SECIS Binding Protein 2 (SBP2) (Caban et al., 2007, Copeland et al., 2001). K-turn structures are flexible and appear to move like a hinge to support their variable conformations (Razga et al., 2005). It has also been reported that certain Mg^{2+} concentrations may inhibit SECIS/SBP2 binding as the SECIS element is fully bent in such conditions (Copeland and Driscoll, 1999).

Another conserved motif of SECIS element is in its apical loop which contains the AAR residue (R denotes purines, A or G) (Cléry et al., 2007). The AAR motif is required for Sec incorporation (Berry et al., 1993), but its mechanism of function remain unknown as no evidence of AAR-binding proteins have been identified to date. Its role in Sec-incorporation is further confounded by SECIS elements found in SelenoM and SelenoO which contains a CC residue in place of the AAR motif (Korotkov et al., 2002, Kryukov et al., 2003).

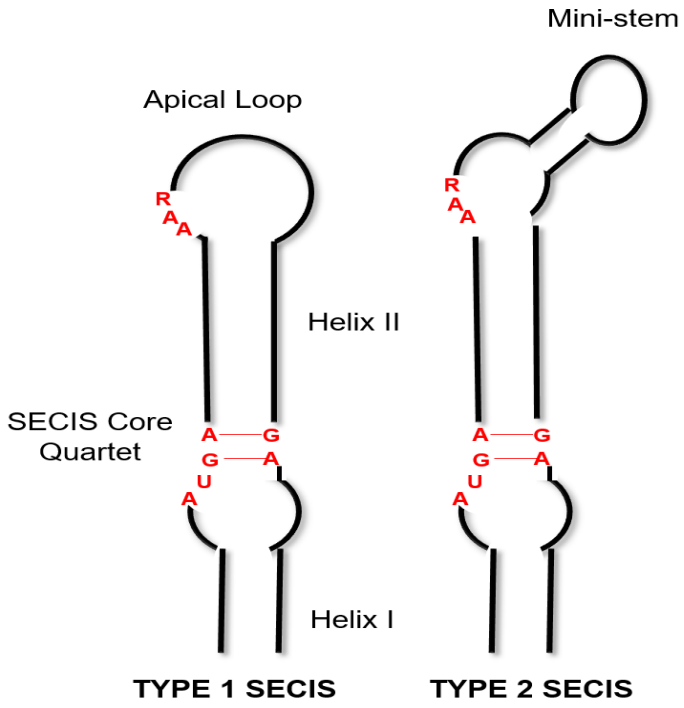


Figure 1.6: Eukaryotic SECIS elements

Type 1 and Type 2 SECIS elements are depicted. Features are indicated as apical loop, SECIS core quartet, helix I, helix II and the mini stem-loop for the type 2 element. The conserved motifs are in red. (Adapted from (Labunskyy et al., 2014).

Selenocysteine Recoding Elements (SRE)

The bacterial SECIS element is located immediately 3' of the UGA codon unlike its eukaryotic counterpart located at the 3'UTR. A homolog of Sec-UGA downstream RNA secondary structure, termed Selenocysteine codon Redefinition Element (SRE) (Howard et al., 2005) was identified in Selenoprotein N (SelenoN). This was found to be functional and forms a stem-loop structure which starts six nucleotides 3' of the

UGA codon. Further experimental analysis (Howard et al., 2007, Howard et al., 2005) illustrated that SRE alone was sufficient to stimulate high levels of UGA readthrough by near-cognate tRNA. The presence of the SECIS element, however, allowed selenocysteine insertion in the absence of the SRE suggesting that the SRE was not required but highly stimulatory. Important features of the SRE that functions in Sec-incorporation includes its upstream sequence, the stem loop and the spacer distance to the UGA codon. An independent genome-wide search of conserved functional RNA structures (Pedersen et al., 2006) independently identified this SRE in Selenoprotein N and Selenoprotein T. More recently, multiple SRE structures were identified in Selenoprotein P (Marrioti et al., 2017), presenting its importance in multiple Sec-UGA decoding (See Section 1.7).

SECIS- Binding Protein 2 (SBP2)

SBP2, the main interacting partner of the SECIS elements, is essential for Sec-UGA recoding during selenoprotein translation (Copeland et al., 2000). It can be divided into three domains: an NH_2 -terminal domain of ~400 amino-acids with no characterized function to date, a central Sec-incorporation domain (SID) consists of ~100 amino-acids and a COOH-terminal domain containing the RNA-binding L7Ae domain (RBD) which binds to K-turn motifs of RNA (Caban et al., 2007, Copeland et al., 2000, Copeland et al., 2001, Allmang et al., 2002, Donovan et al., 2008).

Interestingly, the NH_2–terminal domain is found to be dispensable for Sec-incorporation *in-vitro* while SID and RBD are sufficient for all known SBP2 functions (Copeland et al., 2000, Mehta et al., 2004). The NH_2 domain of SBP2 contains a lysine-rich nuclear localization signal (NLS) and may suggest a regulatory role during nuclear export of the protein (Copeland et al., 2000). The nucleocytoplasmic shuttling of SBP2 appears to be necessary for its role in selenoprotein synthesis as blocked nuclear export resulted to reduced cytoplasmic selenoprotein expression (Papp et al., 2006). Further, nuclear accumulation of SBP2 was observed in cells exposed to oxidative stress suggesting that SBP2 may get shunted to the nucleus, perhaps to avoid oxidative damage (Touat-Hamici et al., 2014, Papp et al., 2006).

The C-terminal domain containing the RBD and SID mediate stable SECIS binding and allow Sec-incorporation (Bubenik and Driscoll, 2007, Takeuchi et al., 2009). Only RBD interacts directly with SECIS while SID is thought to enhance such interactions (Donovan et al., 2008). Despite this, individual SID and RBD proteins can form stable SECIS-dependent complex and retain all their functions in Sec-incorporation *in-vitro* except for stable ribosomal binding (Donovan et al., 2008).

As well as its interaction with SECIS elements, SBP2 has also been shown to interact with eEFSec and the 28S rRNA of the 60S large ribosomal subunit (Reviewed in:(Caban and Copeland, 2006). Such interactions are not well-understood including the actual role SBP2 plays when stably bound to the ribosome. Elaborate structural studies will be required to elucidate these mechanisms.

In addition to SBP2, a paralog of SBP2, named SBP2L has been identified in vertebrates. Despite the sequence similarity between the two versions, mammalian SBP2L does not support Sec incorporation *in-vitro* (Donovan and Copeland, 2009). Functional characterization reveals that SBP2L can bind SECIS but with lower affinity than SBP2 (Donovan and Copeland, 2012). Moreover, in some invertebrates (e.g. the annelid *Capitella teleta*, sea urchins and ascidians) only one copy of SBP2 is present, that displays more similarity to SBP2L which can support Sec-incorporation (Donovan and Copeland, 2012). These observations led to the assumption that SBP2L may play a role in post-transcriptional regulation of selenoprotein expression and may only support Sec incorporation in optimal conditions. More recently, mice lacking *Sbp2* were shown to retain substantial selenoprotein synthesis capacity (Fradejas-Villar et al., 2017, Seeher et al., 2014) suggesting an undetermined role for SBP2L.

Sec-Specific eukaryotic Elongation Factor (eEFSec)

A dedicated elongation factor, eEFSec, is responsible for recruiting Sec-tRNA$^{[ser]sec}$ and in tandem with SBP2, inserts Sec into nascent peptides in response to the UGA codon (Fagegaltier et al., 2000a, Tujebajeva et al., 2000).

During canonical translation elongation, eEF1A is the cellular elongation factor involved in recruiting amino-acylated tRNAs (aa-tRNA) to the ribosomal A-site (Dever and Green, 2012). Similar to eEF1A, eEFSec has GTPase activity,

however, it demonstrates a higher affinity to Sec-tRNA$^{[ser]sec}$ and does not bind other amino-acylated tRNAs (Carlson et al., 2004).

Recent structural analysis reveals similar domain organization between eEFSec and eEF1A (Dobosz-Bartoszek et al., 2016, Gonzalez-Flores et al., 2012). The structure of eEFSec was described as 'chalice-like' and consists of four distinct domains (Böck et al., 1991) with domains I, II and III exhibiting homologies with eEF1A. Like the canonical elongation factor, these domains are involved in tRNA binding, ribosome association and GTP hydrolysis. Domain IV; located at the C-terminal extension and forming the base of the chalice, was uniquely identified for eEFSec. This domain directs the correct delivery of Sec-tRNA to the ribosome when Sec-UGA is present (Itoh et al., 2009a). Moreover, domain IV forms a transient complex with SBP2 (Zavacki et al., 2003). It is also proposed to bind the acceptor-TψC elbow and the variable arm of Sec-tRNA (Dobosz-Bartoszek et al., 2016). A GTP to GDP exchange induces a conformational change in domain IV which might have relevance to the mechanism of cognate tRNA release during Sec-decoding (Dobosz-Bartoszek et al., 2016).

Additional factors

Several canonical translation factors with diverse roles in the cell have also been implicated in Sec-decoding including eukaryotic initiation factor 4a3 (eIF4a3), nucleolin and the ribosomal protein L30 (RPL30).

eIF4a3 is a member of the DEAD-box protein family helicases with characterized roles on RNA metabolism from transcription to degradation (Li et al., 1999b). Notably, eIF4a3 is a member of the Exon-Junction Complex (EJC) which are involved during splicing and eliminates mRNA with pre-mature stop codon via non-sense mediated decay (NMD) (Chan et al., 2004). Despite the in-frame UGA codons in selenoprotein mRNA, only a subset of selenoprotein mRNAs are susceptible targets of NMD to a limited extent including GPX1, GPX4, SELENOH and SELENOW (Sunde and Raines, 2011). It was further noted that eIF4a3 mediates differential binding affinity to SECIS elements where type 1 SECIS elements are preferred (Budiman et al., 2009, Budiman et al., 2011). Such binding inhibits SBP2 interaction by steric hindrance and thus Sec incorporation. In addition, eIF4a3 levels are increased

in selenium deficiency which suggests that eIF4a3 may negatively regulate a subset of selenoproteins in response to selenium level (Budiman et al., 2009).

Nucleolin is a multi-functional protein with critical roles in ribosome biogenesis, transcription and chromatin function (Abdelmohsen and Gorospe, 2012). It selectively binds to a sub-set of SECIS elements *in-vitro* and with high affinity *in-vivo* (Miniard et al., 2010, Squires et al., 2007). Nucleolin knock-down in rat hepatoma cells led to selective downregulation of GPX4 and TXRND1. GPX1 and SELENOF was not affected (Miniard et al., 2010). It was proposed that nucleolin can promote Sec incorporation for some selenoproteins but the mechanism of which remains elusive.

Ribosomal protein L30 is a 14.5 kDa protein that is a component of the large ribosomal subunit (60S) in eukaryotes (Vilardell et al., 2000). It may also exist independent of the ribosome and is involved in other biological processes (Chavatte et al., 2005). Similar to SBP2, eL30 has a K-turn interacting domain (L7Ae) and thus, it binds onto the SECIS element (Xu et al., 2010, Chavatte et al., 2005). RPL30 competes with SBP2 binding to the SECIS element *in-vitro* (Snider et al., 2013, Bifano et al., 2013). RNase footprint assays revealed that SBP2 and RPL30 have some over-lapping and unique individual binding sites on the SECIS element (Bifano et al., 2013). Thus, it was suggested that RPL30 may promote SECIS-SBP2 dissociation allowing for canonical elongation to continue.

Sec-UGA decoding mechanism

Introduction of the eukaryotic Sec-insertion components into a naïve translation system e.g. the wheat germ lysate (WGL) (Shetty et al., 2014) for the translation of selenoproteins has established the minimum core components needed for Sec-UGA decoding. These include SBP2, SECIS elements, eEFSec and the Sec-tRNA. It also appears that non-specialized ribosomes in WGL are competent for Sec-insertion.

A model for Sec-UGA decoding is depicted in Figure 1.7. It is proposed that when the ribosome encounters the Sec-UGA codon, the Sec machinery interacts with the canonical translation machinery to augment the coding potential of Sec-UGA. The SECIS elements at 3'UTR of selenoprotein mRNA governs the recruitment of the Sec-

tRNA, eEFSec and SBP2 (Berry et al., 1991, Bock, 2000, Tujebajeva et al., 2000). SBP2 strongly associates with the ribosome through the L7Ae domain and binds to the SECIS element with high affinity and specificity (Copeland et al., 2000, Copeland et al., 2001, Low et al., 2000). SBP2 also interacts with eEFSec, which recruits Sec-tRNA$^{[ser]sec}$, subsequently incorporating Sec to the nascent peptide (Tujebajeva et al., 2000). RPL30 is proposed to compete with SBP2 for SECIS binding which dissociates the Sec-machinery and allows elongation to continue (Chavatte et al., 2005). Nucleolin and eIF4a3 are regulators of selenoprotein synthesis and may contribute to the hierarchy of selenoprotein expression (Miniard et al., 2010, Budiman et al., 2009).

Recent ribosome profiling analysis of selenoprotein mRNA confirmed the Sec-UGA position as the limiting step of selenoprotein synthesis for most selenoproteins. This was characterized by ribosome accumulation before the UGA codon, a point of regulation possibly needed to recruit the factors necessary for Sec incorporation. Moreover, significant reduction of ribosomes translating downstream of UGA suggests that UGA-decoding is inefficient during selenoprotein translation but can be augmented by Se-bioavailability (Howard et al., 2013).

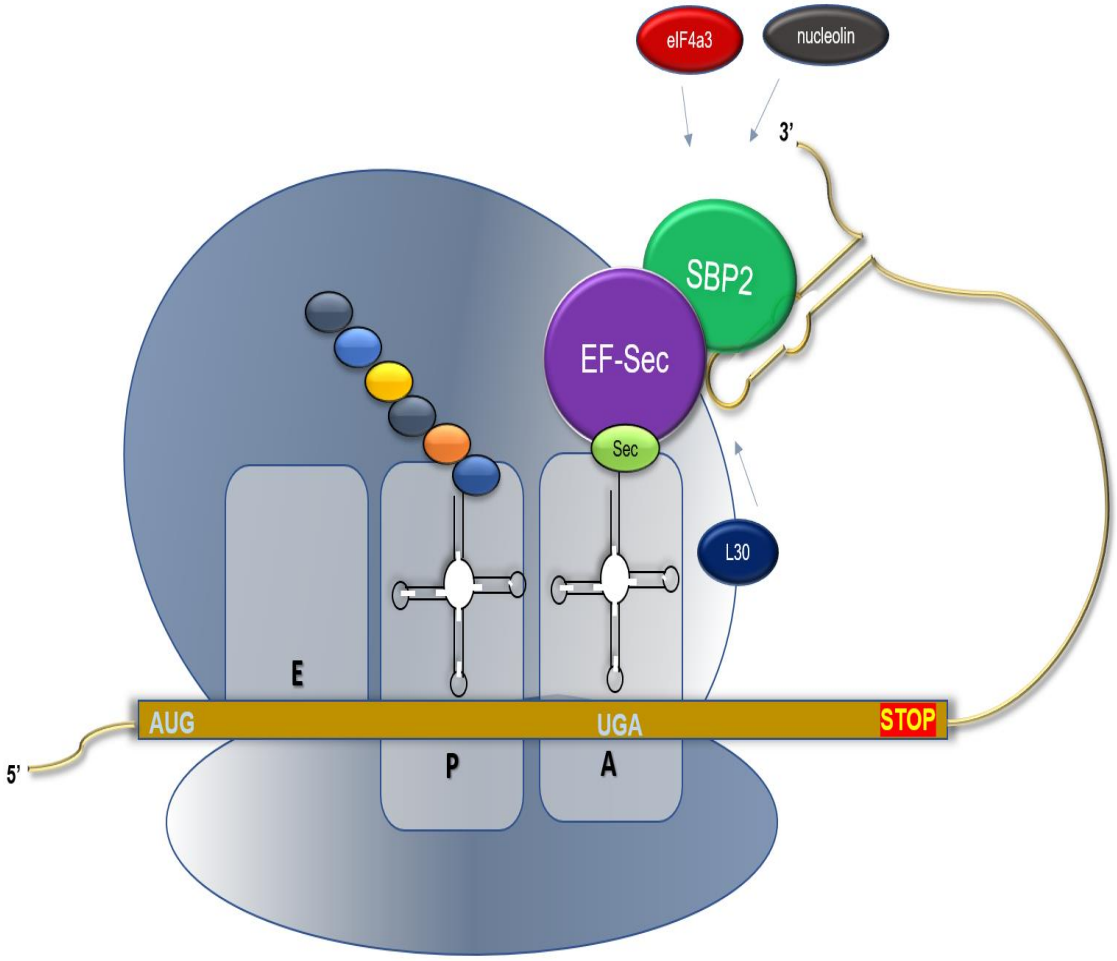


Figure 1.7: Model for Sec-UGA recoding

Known factors involved in Sec insertion during selenoprotein translation are shown including Sec-tRNA, EF-Sec, SBP2, eI4a3, nucleolin and L30 (details are in the text). SECIS interaction with trans-acting factors and Sec-tRNA position on the ribosomal A-site is depicted. (Adapted from (Labunskyy et al., 2014)

1.7: Selenoprotein P (SELENOP) Recoding

As discussed in Section 1.6, the mechanism of a single Sec insertion requires a highly specialized process with concerted interplay of various Sec-dedicated and canonical factors during translation. Selenoprotein P (SELENOP) is unique among selenoproteins in its requirement to redefine multiple UGAs as Sec on its mRNA (between 10 in humans to 28 UGAs depending on species) (Lobanov et al., 2008). We discuss the unique features of SelenoP mRNA and the current proposed model for the mechanism of multiple Sec-incorporation during SELENOP synthesis.

Twenty-five selenoprotein genes have been identified in the human selenoproteome (Kryukov et al., 2003). Most selenoproteins including glutathione peroxidases (GPXs), deiodinases (Dis) and thioredoxin reductases (TRXs) have characterized redox-activity within the cell, therefore function mainly as anti-oxidants (reviewed in (Reeves and Hoffmann, 2009). Such catalytic activity is attributable to the highly reactive Sec-residue present in their active site. In mammals, SELENOP with multiple Sec-UGA, is synthesized in the liver and serves as a selenium transporter across tissues (Hill et al., 2003, Schomburg et al., 2003). Short forms of SELENOP from early termination also exhibits anti-oxidant function (Kurokawa et al., 2014). Further, SELENOP gene inactivation in mice led to both male infertility and neuronal degeneration (Caito et al., 2011, Olson et al., 2005). Despite well-characterized protein functions (Burk and Hill, 2009), the mechanism for SELENOP synthesis remains largely elusive. *In-vitro* and *in-vivo* experiments reveal that a single Sec-incorporation is inefficient (Mehta et al., 2004) which would mean multiple Sec-UGA redefinition is highly unlikely. However, abundant long forms of SELENOP was found in the plasma of rats and human subjects suggesting that SELENOP is efficiently synthesized (Read et al., 1990). Subsequently, reporter construct analysis of SelenoP UGA codons within its native context and inclusion of its 3'UTR demonstrated a highly processive incorporation of multiple Sec residues during synthesis (Fixsen and Howard, 2010).

Consequently, SelenoP mRNA exhibits unique features that may promote processive Sec-incorporation. The first Sec-UGA is isolated in the proximal half of the mRNA while the distal tail contains the rest of the Sec-UGA codons arranged in close-proximity (Hill et al., 1993) . The position of Sec-UGA 1 is also highly

conserved across species (Lobanov et al., 2008). It has been originally proposed that translation efficiency and processivity of SelenoP is governed by SECIS elements on its 3'UTR (Hill et al., 1993) which are also found to be conserved. Unlike single Sec-codon containing selenoproteins, SelenoP mRNA contains two SECIS elements, SECIS 1, a type 2 element and SECIS 2, a type 1 element (Grundner-Culemann et al., 1999).

Earlier functional analysis of SelenoP 3'UTR predicted a model where its SECIS elements have differential functions (Stoytcheva et al., 2006, Fixsen and Howard, 2010). Such model proposes that during initiation, the 3'UTR adopts a loop confirmation and positions SECIS 2 closer to UGA 1. SECIS 2 facilitates decoding of UGA 1 which is shown to be slow and inefficient, leading to a bottle neck in SelenoP translation (Stoytcheva et al., 2006). Once the ribosome translates past UGA 1, SECIS 1 takes over and allows efficient and processive Sec-incorporation in the multiple distal UGAs (Fixsen and Howard, 2010). Ribosome protected fragments mapping to SelenoP mRNA further confirmed the slow decoding of UGA 1 suggested by ribosome accumulation in the upstream region prior to UGA 1 and processive incorporation of the UGA-rich region (Howard et al., 2013). Generation of SelenoP mutant mice with either UGA 1 substitution to serine and SECIS 1 or SECIS 2 deletion had varying effects on the proportions of short or long forms of SELENOP in plasma, and on viability (Wu et al., 2016). Substitution of UGA 1 to serine resulted to reduced long forms of SELENOP in the plasma. SECIS 2 deletion led to the predominant production of long forms while SECIS 1 deletion led to the production of shorter SELENOP forms. SECIS 1 deletion was also lethal in Se-deficient conditions with mice exhibiting neurological damage. These *in-vivo* data further supports the SECIS differential function model and highlights the role of the conserved UGA 1 position in full-length SelenoP translation.

More recent functional studies on SelenoP mRNA revealed coding sequence determinants regulating the efficiency of SELENOP synthesis. A comparative analysis of SELENOP coding regions across species revealed four highly conserved and distinct RNA structures (Mariotti et al., 2017). Two structures, termed Initiation Stem Loop (ISL) and Selenocysteine Recoding Element 1 (SRE 1) are located in the 5' region encoding the N-terminal domain. The region encoding the C-terminal half contains structures named SRE 2 and SRE 3. *In-vitro* functional analysis of ISL

structure, suggests a role at initiation. Mutation of SRE 1, located downstream of UGA 1 reduced Sec-UGA 1 incorporation in cultured cells. The role of SRE 2 and SRE 3 remains to be solved (Mariotti et al., 2017). In addition, sequence context within N-terminal domain were found to be necessary in regulating the efficiency of multiple Sec-recoding in the distal multiple UGAs (Shetty and Copeland, 2018c)

Experimental data from current literature predicts secondary or tertiary interactions within and/or between structures in the coding region and the 3'UTR of SELENOP (Reviewed in (Shetty and Copeland, 2018b). A summary of the current model of multiple Sec codon incorporation during full-length SELENOP synthesis is depicted in Fig. 1.8. The process by which these RNA elements interact with the Sec incorporation machinery or whether additional factors are needed for SelenoP translation warrants further investigation. It is however certain that the features of SelenoP mRNA are unique and required to drive efficient multiple-Sec incorporation events.

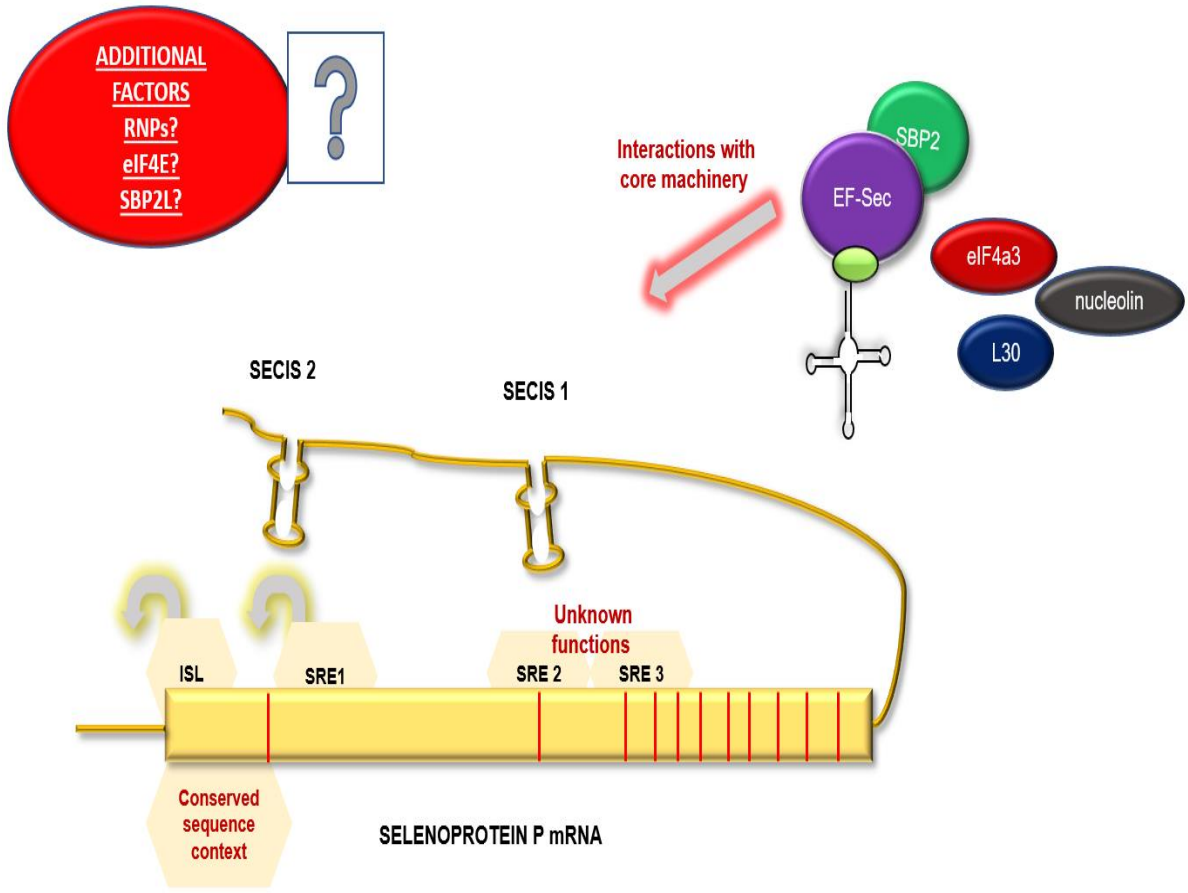


Figure 1.8: Model for selenoprotein P decoding

The figure illustrates the 3'UTR 'looping' model where SECIS 2 facilitates UGA 1 decoding and SECIS 1 facilitates multiple Sec decoding. ISL regulates translation initiation while SRE1 promotes UGA 1 decoding. Function of SREs 2 and 3 are unknown. SELENOP also contains a conserved sequence context at its N-terminal domain. The machinery for Sec-incorporation is also depicted while their interaction with SelenoP mRNA is unknown. Potential involvement of additional factors such as RNPS, eIF4E and SBP2L is illustrated. Red lines depict relative positions of Sec residues. Abbreviations: ISL: Initiation Stem Loop, SRE: Sec recoding elements, eIF4E: eukaryotic initiation factor, RNP: ribo-nucleoproteins, SBP2L: SECIS binding protein 2L. The components of the core Sec-machinery are described in section 1.6.

Chapter 3:

Evolution of Selenoprotein P (SelenoP) in invertebrates with focused characterization of *Crassostrea gigas* SelenoP translation mechanism

Contributions:

Several contents of Chapter 3 are published as part of a manuscript (Baclaocos et al., 2019). To convey a complete description of the work undertaken, the following contributions from collaborators were included:

- Dr. Marco Mariotti and Dr. Didac Santesmasses performed the phylogenetic analysis, evolutionary reconstruction and structure predictions. They also generated Table 3.3.1, Figures 3.3.1, 3.3.2 and 3.3.12.
- Radioactive ^{75}Se experiments were performed by the thesis author in collaboration with Dr. Paul Copeland et al., in Rutgers New Jersey, USA.
- All other experimental work and analysis in this Chapter was performed by the thesis author, unless otherwise stated.

3.1 Introduction

3.1.1: Background

Selenium (Se) utilization is known to occur across species in three different ways. It is present in the form of 2-selenouridine located in the wobble position of certain bacterial tRNAs (Wolfe et al., 2004). It also occurs as a labile co-factor of molybdenum (Mo) containing proteins in bacteria where a Se-Mo moiety is directly involved in catalysis (Dilworth, 1982, Gladyshev et al., 1994, Self and Stadtman, 2000). Perhaps the most studied form of Se is one that is incorporated as the 21st amino-acid selenocysteine (Sec) into selenoproteins (Labunskyy et al., 2014, Hatfield et al., 2014). The majority of selenoproteins exhibit an oxidoreductase function with Sec typically located in the active site similar to thiol-based oxidoreductases (Kim and Gladyshev, 2005). With a small number of interesting bacterial exceptions (Vargas-Rodriguez et al., 2018, Mukai et al., 2016), selenocysteine is encoded by UGA (Zinoni et al., 1987). While in standard decoding UGA specifies translation termination, in response to mRNA-specific recoding signals and multiple specialized accessory factors, its meaning is dynamically redefined to specify selenocysteine in selenoproteins as discussed in Chapter 1, section 1.6.

Selenoprotein annotation is often problematic as the in-frame UGA codon is usually interpreted as a translation-termination signal. However, several bioinformatic tools have been developed to identify selenoprotein sequences. The original approach focused on Selenocysteine Insertion Sequence (SECIS) element identification. Eukaryotic SECIS elements are found in the 3'UTR of selenoprotein mRNA and contain conserved sequences forming a unique and stable RNA secondary structure (see section 1.6). The first selenoproteins identified by virtue of identification of such SECIS elements within their mRNA sequences were mammalian selenoproteins R, N and T (Lescure et al., 1999, Kryukov et al., 2003). Analysis of closely related genomes allowed for identification of conserved SECIS elements belonging to selenoprotein orthologs in these organisms (Kryukov et al., 2003). These techniques were further refined and have led to the development of bio-informatic tools for SECIS element and selenoprotein identification, most notably, SECISSearch3 (Mariotti et al., 2013), Selenoprofiles (Mariotti and Guigo, 2010) and Seblastian (Mariotti et al., 2013). Independent of SECIS identification is another approach that searches in-frame UGA

codons flanked by conserved sequences in selenoprotein coding regions (Kryukov et al., 2003, Kryukov and Gladyshev, 2004, Castellano et al., 2001, Zhang et al., 2005b). Such conserved regions are also present in selenoprotein homologs with cysteine (Cys) instead of Sec. The Sec/Cys homology and SECIS-based algorithms identify very similar sets of selenoproteins independently and are often used together to annotate the selenoproteomes of newly sequenced genomes in order to advance selenium or selenoprotein studies. Experimental confirmation via radioactive ^{75}Se labelling is usually performed to establish metabolic Sec incorporation into selenoproteins (Yim et al., 2018).

Analysis of the selenoproteome has been used to uncover trends in Se/Sec utilization across species (Lobanov et al., 2007) like massive independent selenoprotein losses in some insects, higher plants, fungi and protists (Mariotti et al., 2015, Chapple and Guigo, 2008). Evolutionary reconstruction of selenoproteomes points to selenoproteins having ancestral origins as such genes can be traced back to unicellular eukaryotes (Lobanov et al., 2007). In addition, aquatic organisms have been shown to contain an expanded selenoproteomes in contrast to terrestrial organisms, which have reduced their selenoproteomes through loss of selenoprotein genes or Sec to Cys codon replacement (Lobanov et al., 2008). The current hypothesis to explain these observations is an increased Se-requirement in aquatic environments and a reduced Se-reliance in terrestrial habitats (Lobanov et al., 2007) suggesting that the environment may play a role in selenoproteome evolution.

Nearly all selenoproteins contain a single selenocysteine which is at the active site. Selenoprotein P, SELENOP (Gladyshev et al., 2016) contains a single N-terminal domain Sec, which has a proposed redox function (Kurokawa et al., 2014) in addition to multiple selenocysteines in its C-terminal (9 in humans and rodents) (Hill et al., 1991, Himeno et al., 1996, Ma et al., 2002). The long isoforms serve a selenium transport function in mammals where delivery is critical to the brain, testes and other tissues (for review, please see (Burk and Hill, 2015). Known SelenoP mRNAs are also unique among selenoprotein mRNAs in having 2 rather than 1 SECIS elements in their 3' UTRs (Berry et al., 1993). In addition, secondary structures within mRNA coding sequence termed Sec-Recoding Elements (SREs) (Howard et al., 2005) and Initiation Stem Loop (ISL) were identified (Mariotti et al., 2017). These RNA structures have

proposed roles in enhancing UGA-decoding efficiency and translation initiation of mammalian SelenoP *in-vitro*.

Apart from limited studies in amphioxus and sea urchins (Lobanov et al., 2007), SelenoP has been almost exclusively investigated to date in vertebrates. Reconstruction of evolutionary changes in the Se-transport domain of SELENOP in vertebrates revealed a decrease in Sec content, specifically in the mammalian lineage via Sec to Cys replacement (Lobanov et al., 2008). In comparison to mammals, bony fishes demonstrated higher Sec content, larger selenoproteome, elevated *selenoP* gene expression and higher tissue Se. Therefore, SelenoP evolution from fish to mammals accompanied by Sec reduction, smaller selenoproteomes and Sec to Cys transition events has been proposed as a genetic marker of Se utilization in vertebrates.

3.1.2: Chapter Aims

Selenoprotein P (SelenoP) is a special case among all other known selenoproteins as the production of the full-length protein necessitates redefinition of multiple UGA codons to selenocysteines as well as the possession of two SECIS elements with proposed differing functions (Ma et al., 2002). Bioinformatics analysis of vertebrate SelenoP has shown up to 22 in-frame UGA codons (Mariotti et al., 2012) while the largest number of UGAs identified is in an invertebrate, the sea urchin, where 28 are present (Lobanov et al., 2008).

In this chapter, an extended search of invertebrate SelenoP sequences will be described in order to gain insights to its evolutionary history. Such information will be useful to improve our understanding of how the mechanisms of translational processivity has evolved. Further, to investigate the genetic-decoding versatility of SelenoP translation across species, representative organisms were selected to test their translation in an established supplemented heterologous system. Since an unprecedentedly high numbers of Sec-encoding UGAs was identified specifically in molluscs SelenoP, a focused characterization of *Crassostrea gigas* (*C.gigas*), Pacific oyster, SelenoP mRNA translation and Sec-insertion machinery components was performed. The origin of Sec-rich extension in SelenoP and the diversity of Se/Sec utilization will be discussed as well as some apparent differences in the mechanism of Sec-UGA decoding identified in the oyster SelenoP compared to its mammalian counterpart.

3.2 Materials and Methods

3.2.1 SelenoP gene finding, phylogeny, filtering

Gene prediction was carried out with the program Selenoprofiles v.3.5c (Mariotti and Guigo, 2010), which employs a protein alignment profile to scan nucleotide databases and find genes belonging to the same family. Due to its peculiar C-terminal Sec-rich domain, SelenoP is particularly difficult to predict. This domain contains various stretches of repetitive sequences, resulting in lots of spurious hits in non-homologous repetitive regions of genomes during the very first step of Selenoprofiles (blast search). These hits can be easily recognized upon manual inspection, but due to the magnitude of sequences for analyses a new version of selenoprofiles with various filtering parameters was produced *(manuscript submitted*). These parameters exclude poor quality sequences, incomplete gene structures and predictions from putative contamination. SelenoP from vertebrates formed a clear monophyletic cluster; thus, this portion of the tree was separated from the rest of the metazoan sequences for visualization purposes and will not be discussed further in the results section. The final bona-fide set consisted of 1,228 SelenoP predictions. We employed this tree to identify several protein clusters, represented in Fig. 3.3.1. Multiple sequence alignments of representative sequences from particular taxonomic groups (e.g. gastropods or cephalophods) were carried out using ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Larkin et al., 2007) and images were generated using a downloaded version of Jalview 2.10.5 (Waterhouse et al., 2009).

3.2.2 SelenoP plasmid construction and oyster SelenoP mutant construct generation

Native cDNA sequences were used to generate gene blocks that were obtained from Integrated DNA Technologies (IDT), one candidate sequence (from the spider *Parasteatoda*) was selected for gene synthesis (GenScript) due to sequence complexity for gene block generation. C.*gigas* selenoP gene was amplified from genomic DNA to include intronic sequences. All SelenoP gene-block constructs were cloned into the pcDNA3.1 TOPO-TA vector by BamHI and AgeI-HF digestion.

A plasmid encoding zebrafish selenoP positive control construct was a gift from Paul Copeland's lab (Shetty and Copeland, 2018c). Mutant constructs were

generated by either PCR or site directed mutagenesis. Kozak context (construct O2) was added by Quick Change II kit (Agilent Technologies) site directed mutagenesis following manufacturer's protocol and with primer sequences indicated. O3 was PCR amplified with primers indicated and cloned into double-digested pCDNA3.1 TOPO-TA vector. Synonymous mutation of the ISL (O4) was performed by a 3-step PCR introducing synonymous codons at each PCR step. Removal of the distal region with multiple UGAs (O5) was carried out by a two-step cloning procedure where CDS was re-ligated with the 3'UTR after UGA 11-46 removal by PCR. Addition of oyster SECIS elements to zebrafish coding region was done by Pac1 and Not1 double digest removal of zebrafish SECIS 1 and re-ligation of oyster SECIS amplified by PCR including restriction sites. Oyster SECIS elements were subcloned into the pcDNA3.1 TOPO-TA vector containing zebrafish CDS insert (O7, O8, O9). All restriction enzymes used for each insert are indicated in Table 3.2.2. Inserts digested with BglII/Age1-HF were inserted into a BamH1/Age1HF cut pCDNA 3.1 TOPO-TA vector (a gift from Paul Copeland's lab) while the same restriction enzymes were used to cut the vector for all the other mutants.

Table 3.2.1: Table of primers used for mutant oyster constructs cloning

Construct	Forward (5'-3')	Reverse (5'-3')	Restriction enzymes
O1 Native	ATAAAAGATCTAAATCACGTG	TTATACCGGTATGAATATTGATCA	BglII/ AgeI-HF
O2 Kozak consensus added	GACAAAGCTCCGGACCATGGCAATGCGGGGCCCCG	CGGGGCCCCGCATTGCCATGGTCCGGAGCTTTGTC	BglII/ Age1-HF
O3 5'UTR removed	ATAAGA TCTAGAACCATGGGGCGGGGCCCCGCCGGCCTTTG	CTTATTACCGGTATGAATATTGATCATACTGAAG	Xba1/ Age1-HF
O4 Initiation Stem Loop (ISL) Synonymous codon Disruption	AGT AGC GCA TTG GGG CAG ACA TGT CAA CGT AGC ACT TTA CCA TGG CGA ACA GCC GAC GGA *(PCR1, 10cycles, change positions 90-120)*	TTATACCGGTATGAATATTGATCA	Xba1/ Age1-HF
	TTTGGCTGGTCGCGTGGCTGAC A GCT GCT GTT CTA AGT AGC GCA TTG GGG CAG ACA *(PCR2, 10cycles, change positions 40-89)*		
	ATAAGA TCTAGA **ACC ATG GGG** CGGGGCCCCGCCGGCCTTTGGC TGGTCGCGTGGCTG *(PCR3, 30cycles, add start sites and restriction sites)*		
O5 UGA 11-46 removal	ATAATAGCGGCCGCATGCGGGGCCCCGCCGGCCTTTGGCTGG *(Removal of Sec-rich domain)*	TTCCAACTCTAGATTACTAGGATCCTCGAGACCCTTGTCACT	Not1-HF/ Xba1
	ATAATATCTAGA AGAGTAGGGTCTCGTCAAATCG TCTCTACTGGTAA	TGCTACTTTTTTTCTTCAGTATGTCAATATTCATACCGGTTTAAAGA	Xba1/ Age1-HF

	(3'UTR amplification)		
O6 Zebrafish fusion with oyster 3'UTR	GGAATATTAATTAATAGAGAGTAGGGTCTCGTC	GCTAAAATGCGGCCGCATGAATATTGATCATACTGAAGAA	Pac1/ Not1-HF
O7 Zebrafish fusion with oyster SECIS 1	GGAATATTAATTAATGTTGTATAGAGCATCTGTAATCAGTGCCATGAAG	GCTAAAATGCGGCCGC GAGTATACAATTCTTTACCGTGTTAATCATTAA	Pac1/ Not1-HF
O8 Zebrafish fusion with oyster SECIS 2	GGATAGTT TTAATTAAAAGTAGACTCTATATCATGTCATCATATTTTCAGAAA	GCTAAAAT GCGGCCGC GAGTATACAATTCTTTACCGTGTTAATCATTAA	Pac1/ Not1-HF

After the double digest step, a 10 µl ligation reaction was prepared with 3:1 insert to vector dilution, 1µl 10X ligation buffer (NEB) and 1µl T4 DNA ligase (NEB). These were left overnight at 4°C. Transformation was carried out in 45µl of competent DH5α cells by incubation with each ligation reactions on ice for 20 min, followed by heat-shock at 42°C for 90 s and back again on ice for 2 min. The DH5α transformed cells were recovered by 200µl LB-media addition and incubation at 37°C for 30 min. The cells were plated in LB-AMP plates and incubated at 37°C overnight. The following day, 10 colonies were selected, lysed by boiling with 10µl of H_2O at 95°C for 5 mins and subjected to colony PCR using FirePol polymerase (Solis Biodyne) as per the manufacturer's guideline. Specific primers (the same as listed in Table 3.2.1) were used to identify positive clones. Plasmid DNA was purified using GeneJET Mini-Prep Plasmid Isolation Kit (Thermo Fisher) as per the manufacturer's guideline. Purified DNA was sent for a commercial Sanger sequencing service offered by Eurofins (www.eurofinsgenomics.eu) to confirm correct sequence insert.

3.2.3 Oyster SBP2 recombinant protein purification

Oyster SBP2 was obtained as a gene-block sequence and codon-optimized for bacterial expression. Full-length oyster SBP2 coding sequence was digested by *Xba1* and *Sac1HF* and ligated into a home-made bacterial vector (Pj307) (Antonov et al., 2013) cut with *Spe1* and *Sac1HF*. Subsequent cloning procedures were carried out similarly to those described in section 3.2.2.

The empty vector encodes an N-terminal Glutathione Sepharose Transferase (GST) tag followed by an HRV-3C protease cleavage site, a 6XHis-tag at the C-terminal end. A positive clone with the SBP2 insert was transformed into BL21 competent cells. A colony was selected and grown in 5ml of LB-AMP media at 37°C overnight. Bacterial cells were then transferred to 3L of LB-AMP media and grown at 37°C for 2h and induced with 0.05 mM IPTG overnight at 15°C once it reached an OD_{600} of 0.6.

Bacterial cells were harvested by centrifugation at 5000 x g for 10 min and washed once with PBS. The cells were lysed in ice cold lysis buffer (1X PBS pH7.2, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, 0.1% NP-40, 100 µg/ml lysozyme and protease inhibitors). The lysates were incubated in 4°C to allow for cell-wall rupture by lysozyme and sonicated three times on ice with a single 30 second burst (probe amplitude 12µm). Resulting lysates were centrifuged at 12,000 x g for 30 min and were incubated overnight with glutathione sepharose beads (GE Healthcare Sigma 17-0756-01) at 4°C. After binding, GST beads were washed three times with wash buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 0.1% NP40). A final wash was performed with 10 bead volume of protease cleavage buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 0.1% NP-40, 1 mM DTT and 1 mM EDTA). GST beads were then resuspended in 1 bead volume of cleavage buffer and 40 µl Prescission Protease to cleave off the GST-tag (GE Healthcare, 27-0843-01). Resulting cleaved protein was concentrated in Amicon ultra-centrifugal filter units with 30kDa molecular weight cut off (MWCO) concentrators (Sigma). The concentrated preparation was subjected to gel filtration using Superdex-200 column and a subsequent buffer exchange to PBS was performed. Protein preparations were analysed by SDS-PAGE electrophoresis followed by Coomassie gel staining and a parallel Western blot.

3.2.4 Rabbit Reticulocyte Lysate (RRL) translation and ^{75}Se labelling

Plasmid DNAs were isolated from positive clones (Promega) determined from colony PCR and sequencing. Purified plasmid DNA were used to generate capped mRNA following manufacturer's protocol (mMessege mMachine Ambion) for *in-vitro* expression of constructs in rabbit reticulocyte lysate (RRL) reactions (Rabbit Reticulocyte Lysate System, Nuclease Treated L4960 Promega). RRL reactions were reconstituted with rat CT-SBP2 or oyster SBP2 and radiolabelled with ^{75}Se as previously described (Shetty et al., 2014). RRL reactions were denatured with protein denaturing sample buffer and boiled at 95^{0}C for 5 min. Samples were resolved in 10% SDS-page gel, fixed, dried and exposed for radio-labelled product detection by PhosphorImager analysis (GE Healthcare).

3.2.5 Cell culture transfection and ^{75}Se labelling

HEK293T cells cultured in Hyclone EMEM (Thermo Fisher Scientific) media containing 10% FBS (Thermo Fisher Scientific) and maintained at 37°C with 5% CO_2, were used to *in-vivo* detect SELENOP products of the invertebrate SelenoP constructs. For transfection, jetPRIME (Polyplus) reagent was used according to manufacturer's protocol. Cells were seeded at 5-6 x 10^5 density per well of a 6-well plate 24h prior to transfection. 24h post-transfection, serum-free EMEM supplemented with 100 nM of ^{75}Se (specific activity of 6.29 μCi/μl; Research Reactor Center, University of Missouri, Columbia). The media was collected and centrifuged the following day at 2500 x g for 5 min at 4°C. The top 80 % of the centrifuged media was transferred to a new tube. 1.5% of total centrifuged media was used for analysis by 10% SDS-PAGE. The adhered cells were then gently washed with cold PBS and lysed (50 mM Tris-HCl pH 8.0, 150 mM sodium chloride, 1% NP-40, Roche complete Protease Inhibitor). The pre-cleared lysate was analysed using PhosphoImager Analysis described in 3.2.4.

3.2.6 Reverse Transcription-Polymerase Chain Reaction (RT-PCR) and Sanger-sequencing

To sequence SelenoP from a fresh-water mussel, *Elliptio complanata (E.complanata)* included in our study, we obtained native *E.complanata* material (whole animal) from Department of Natural Resources, Maryland, USA (Department of Natural Resources

580 Taylor Ave., C-2 Annapolis, Maryland 21401). Total RNA was isolated from the *Elliptio* tissue using Trizol according to manufacturer's instructions (Invitrogen™ 15596018). Reverse transcription (RT) was performed with polyA oligo dT primer using SuperScript® III First-Strand Synthesis System for reverse transcription (Invitrogen 18080051) according to the manufacturer's manual. Polymerase chain reaction (PCR) was carried out using Phusion polymerase (NEB) as per manufacturer's protocol using the forward primers 5' TAAACTGGCTGAGCCGCGGTGGCTC 3' (T1) and 5' AAGTGACAAAACGCACAAGCAATGTTAAAATGCAC 3' (T2) and reverse primer reverse primers 5' ATATATAATATGCTAAGAGTGATCA 3' (for T1 or T2). Amplified PCR product was gel purified using a Zymoresearch DNA isolation kit as per the manual and sent for Sanger sequencing using the following forward sequencing primers:

5' TTGCATTCCCTGCTTTTGACATACTGTCG 3' (S1),

5' ACGTTGATTTTTGATGACGTCACTCAGAT 3' (S2) and

5' TCCATTTGATGAAAGAAAGCTAAAGAA3' (S3).

3.2.7 Structure prediction

RNA secondary structures were predicted using SECISSearch3 or RNAfold webserver (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi) and images were downloaded from this website.

3.3 Results

The advent of next-generation sequencing technologies has led to a tremendous expansion in our genomics knowledge across species revealing unprecedented insights on the evolution of specific genes or proteins. The abundant availability of new sequences was exploited to examine SelenoP evolution in metazoans, revealing a remarkable diversity outside vertebrates.

3.3.1: Characterization of SelenoP across invertebrates

The search for SelenoP coding sequences in metazoan genomes and transcriptomes resulted in 3,464 predictions. In order to obtain a bona-fide set for a phylogenetic reconstruction, these were processed to remove low quality and redundant sequences. A conservative 'phylogenetic filter' (Baclaocos et al., 2019) based on the principle that evolutionary events (protein elongations, deviations from species topology to protein tree) were only trusted if observed across multiple species. The final set contained 1,288 SelenoP sequences found in ~50% of the metazoan genomes searched, clustered in specific lineages. SelenoP was not identified from the genomes of tunicates, Platyhelminthes, all nematodes except *Trichinella*, and the great majority of arthropods (Table 3.3.1). From this analysis, presence of UGA rich distal segments was apparent in the SelenoP of various metazoan lineages. Diverse 3' region extensions were observed in echinoderms, arachnida, cnidaria, and various lophotrochozoa (lineage including annelida, nemertea, molluscs). To investigate the evolution of SELENOP and the emergence of the Sec-rich C-terminal domains, a gene tree was reconstructed using the N-terminal portion of an alignment of all the filtered SELENOP proteins. The resulting invertebrate phylogenetic tree, pruned to remove filtered out predictions, is presented in Fig 3.3.1. Next, prominent features of SelenoP in the various metazoan lineages are described (vertebrates are omitted in the following description).

Cluster	Description	Colour Code
echinoderm	Multi-Sec SelenoP in echinoderms (e.g. purple sea urchin)	Chocolate
Insect	SelenoP with single Sec or Cys instead in basal insects (e.g. dragonfly, cockroach)	Orchid
arachnida	Multi-Sec SelenoP in certain arachnida (e.g. common house spider)	Plum
arachnida2	Multi-Sec or single-Sec SelenoP in certain arachnida (e.g. deer tick)	SlateBlue
arachnida3	SelenoP in certain arachnida (e.g. chigger mite); with single-Sec or diverse substitutions	CadetBlue
Mollusca	Multi-Sec SelenoP in mollusca, including bivalves (e.g. oyster), gastropods, cephalopods	Sienna
gastropodC	SelenoP in certain gastropods such as Spanish slug; converted first Sec to Cys	GoldenRod
Worms	Multi-Sec SelenoP in Nemertea (ribbon worms; e.g. bootlace worm) and annelids	Aquamarine
Cnidaria	Multi-Sec or single-Sec or Cys SelenoP in certain cnidaria (e.g. stalked jellyfish)	SteelBlue
cnidaria2	Multi-Sec in certain cnidaria (e.g. moon jelly); converted first Sec to Cys	MediumBlue

Table 3.3.1: Phylogenetic protein clusters.
The table contains a list of groups of proteins, each forming a monophyletic cluster in the SelenoP gene tree based on aligned N-terminal sequences. These clusters were used to colour-code genes presented in Figure 3.3.1.

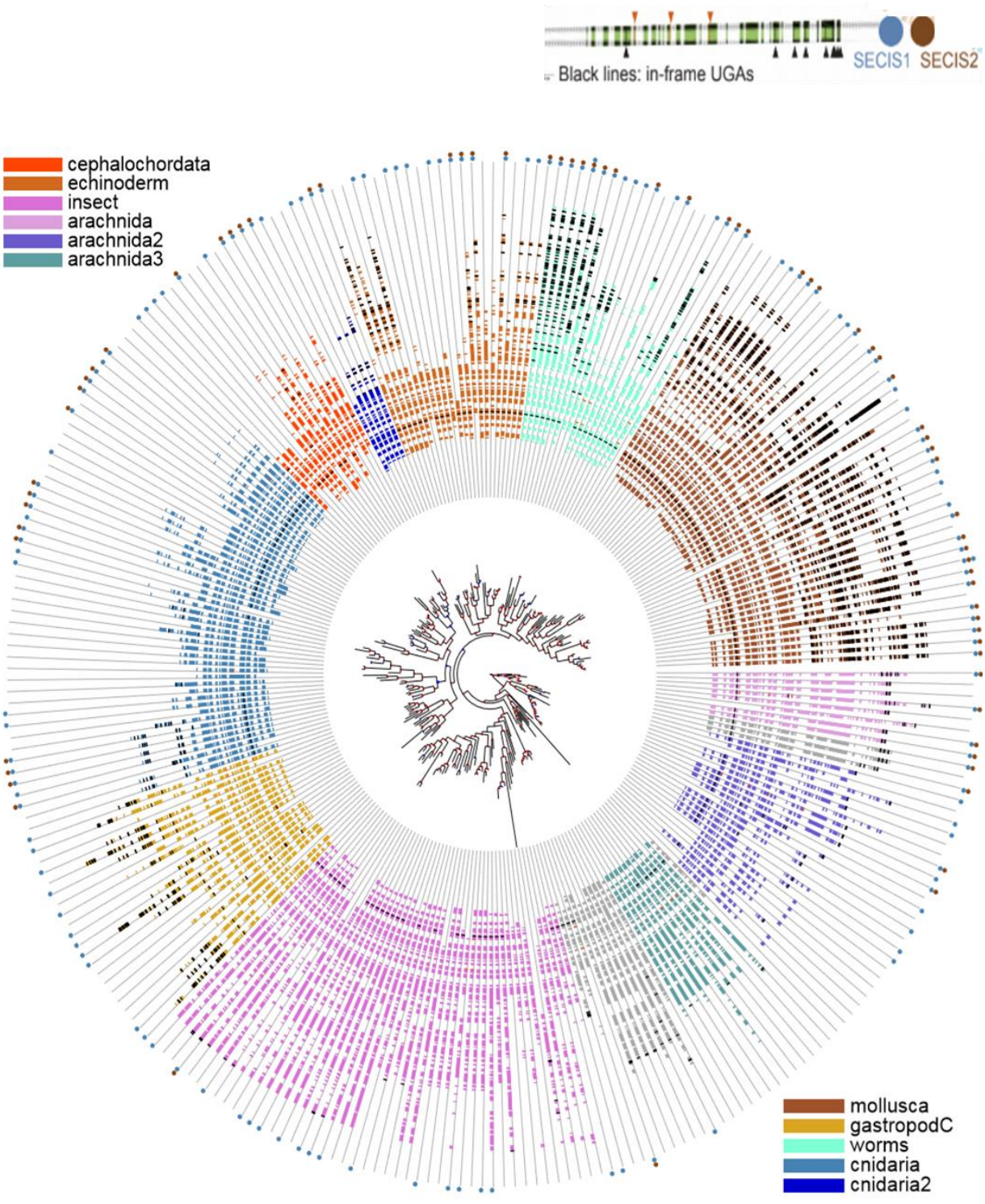


Figure 3.3.1 (electronic link for better resolution)

Figure 3.3.1: Phylogenetic tree of invertebrates with SelenoP genes.
Simplified phylogenetic tree of SelenoP genes in invertebrates. This representation displays the positions of putative Sec-UGAs as black lines while the coloured regions are protein coding sequences. Genes are coloured by protein clusters (see Fig 3.3.1) derived from the gene tree topology based on aligned N-terminal sequence homology. Next to each gene, predicted SECIS elements are shown in blue (SECIS 1) and red (SECIS 2) circles.

Cephalochordates

The selenoproteome of cephalochordate *Branchiostoma floridae* (amphioxus) has been previously reported (Jiang et al., 2012). SelenoP is found here with multiple Sec-encoding residues (UGAs) and two SECIS elements. However, instead of having a UGA-rich distal region like other species, amphioxus SelenoP is formed by tandem repetitions of sequences encoding the thioredoxin-like domain, so that each of its Sec codon is in a homologous context to the characteristic UGA 1 position of human SelenoP. Through the genomic searches, this gene structure was found to be conserved in the cephalochordate lineage.

Echinoderms

SelenoP was previously reported in the echinoderm *S. purpuratus* (purple sea urchin) (Lobanov et al., 2008). Indeed, the searches identified SelenoP in all echinoderms, forming a single monophyletic cluster in the reconstructed tree. SelenoP contains multiple Sec-UGAs and two SECISes in many species besides *S. purpuratus*, with the maximum number of Sec codons (43 UGAs) identified in brittlestar *Amphiura filiformis*. However, several other species (e.g. sea cucumber *Parastichopus parvimensis*) contained instead, a SelenoP gene with single SECIS and Sec, at UGA 1.

Other Lophotrochozoa

The Lophotrochozoan lineage includes Molluscs (discussed separately), Platyhelminthes (flatworms), Nemertea (ribbon worms) and Annelida (segmented worms). While SelenoP-like sequences was not found in Platyhelminthes, the other two categories of worms presented highly UGA-rich SelenoP genes. For example, 65 in-frame UGAs are present in SelenoP of bootlace worm *Lineus longissimus* (Nemertea); 66 are found in *Platynereis dumerilii* (Annelida). Two SECISes were identified in most multi-Sec SelenoP in worms. In the reconstructed gene tree, worm sequences formed a single cluster that also included brachiopods (e.g. *Lingula anatina*), sister to mollusc SelenoP, consistent with taxonomy. This cluster includes a few genes with a single UGA and one or no SECIS (although this could be due to poor assembly quality). The region encoding the C-terminal Sec-rich domain has evident

homology between Mollusca, Nemertea and Annelida, supporting common inheritance of Sec-rich SELENOP within Lophotrochozoa.

Nematodes

SelenoP was absent in the great majority of nematodes, with the only exception of the *Trichinella* genus (early-branching lineage of parasites). Here, distant SelenoP homolog in several species was identified, carrying a Cys codon at the characteristic UGA1 site. While in some cases there are 2-3 in-frame UGAs at the end of the CDS, these UGAs are not conserved in this genus, and no SECIS candidate is present for any of these genes except one, wherein it is located rather distant from the CDS (~2.5kb), strongly indicating that these UGAs do not code for Sec. Thus, the sequence found here is believed not to be a selenoprotein in *Trichinella*, and that other nematodes either lost this gene entirely or it diverged beyond bioinformatic recognition.

Insects

SelenoP was absent in the great majority of insects, including fruit flies, mosquitos and beetles. However, SelenoP was identified in various early branching insect lineages, including Zygentoma (silverfish, firebrats), Odonata (dragonflies, damselflies), Blattodea (cockroaches, termites), Phasmatodea (stick-bugs), and Paraneoptera (lice), forming a single protein cluster in phylogenetic reconstruction. In these species, SelenoP is typically found with a single SECIS and Sec codon at UGA1 position. Beyond the region encoding for the N-terminal thioredoxin-like domain, insect SelenoP genes contain a region encoding ~550 residue C-terminal extension with no similarity to any characterized protein. All the above-mentioned insect lineages contain the Sec-machinery, which is not ubiquitous in this class: several branches of holometabolous insects (Endopterygota) lost it in independent events (Chapple and Guigo, 2008, Mariotti et al., 2015). Consistently, SelenoP was absent in holometabola, with the only exception of some Hymenoptera. In these selenoproteinless species, UGA1 is not conserved and there is no detectable SECIS, indicating that this remote hymenopteran SelenoP homolog is not a selenoprotein.

Other arthropods

A remarkable diversity of *selenoP* genes was found in arthropods. Crustaceans appear not to contain SelenoP. The centipede *Strigamia maritima* contained instead a gene with single a Sec at UGA1. Within arachnida, the sequences formed three phylogenetic clusters located in basal position in the reconstructed protein tree. Despite this topology, there is clear homology between the region encoding the C-terminal domains of all these clusters, pointing to a single orthologous group with diverse divergence rates. The first cluster consisted of sequences from Araneae (spiders, such as *Parasteatoda tepidariorum*). These genes carry multiple Sec codons and two SECIS elements. The first Sec codon is at the characteristic UGA1 site, while the rest are in a C-terminal extension featuring obvious modularity (detailed in Discussion). The second cluster consisted of sequences from Ixodoidea (ticks, such as *Ixodes scapularis*). These genes have a single Sec codon located at UGA 1 position. SECIS was identified in some, but not all sequences, possibly due to incomplete transcripts or high SECIS divergence. The third arachnid SelenoP cluster consisted of Acariformes (mites, such as *Leptotrombidium deliense*). This group included both genes with an in-frame UGA corresponding to the UGA1 site, and genes with diverse codons in its place. Surprisingly, no SECIS elements were detected elements in this group.

Cnidaria

Occurrences of SelenoP were detected in many Cnidaria (early branching lineage of marine metazoans). Upon protein tree reconstruction, cnidarian sequences formed two distinct phylogenetic clusters. The smallest group consisted of sequences from Scyphozoa (true jellyfish such as *Aurelia aurita*, which has paralogs in both clusters). Genes in this cluster have a distal 3' tail with multiple UGA codons encoding for Sec, but UGA1 is replaced with a Cys codon, and only one SECIS element was identified. The other cnidarian phylogenetic cluster encompassed all Cnidaria (hydrozoans, box jellyfish, sea anemones, corals). In various species, this gene presented a region encoding a Sec-rich C-terminal domain with clear similarity to the one of Scyphozoa (indicating homology between the two clusters), but also has a Sec codon at UGA1 and two SECIS elements. The majority of Cnidaria, however, encodes a single-Sec thioredoxin-only SelenoP, which surprisingly featured two SECISes in many cases.

Molluscs

SelenoP sequences were identified in all main classes of molluscs: bivalves, cephalopods, and gastropods. With few exceptions ascribed to imperfect assembly quality, all molluscan SelenoP genes encode a long Sec-rich C-terminal domain, with clear homology within molluscs. Here, the highest number of putative Sec codons in any species was identified in the freshwater mussel *Elliptio complanata*. While *E. complanata* constitutes an outstanding outlier, high Sec counts were present in most bivalves (e.g. 46 in Pacific oyster *C. gigas*), and many cephalopods (e.g. 66 in golden cuttlefish *Sepia esculenta*) and gastropods (e.g. 45 in veined rapa whelk *Rapana venosa*; 14 in common periwinkle *Littorina littorea*). Within gastropods, the Sec codon at the characteristic UGA1 position was replaced by a Cys codon in the lineage comprising of Heterobranchia and Caenogastropoda (Fig 3.3.2). The translated SelenoP mRNA form a separate protein cluster in the gene tree, suggestive of increased divergence rate. While two SECISes were identified in many other mollusc-multi-Sec codon SelenoP, only one SECIS was identified in this cluster. In a few cases (e.g. genus *Crepidula*), no SECIS was identified, and the CDS did not carry any conserved in-frame UGAs, suggesting that SelenoP lost all Sec residues in these species. The phylogenetic cluster includes a single gene with Sec codon at UGA1, from owl limpet (*Lottia gigantea*; Patellogastropoda) and only one SECIS element. While the phylogenetic relationships of gastropods are still uncertain, the placement of this Patellogastropoda gene with Heterobranchia and Caenogastropoda apparently contradicts the most widely accepted taxonomy, as Patellogastropoda is considered an early-branching gastropod (Zapata et al., 2014). This suggests a complex gene history for gastropod SelenoP.

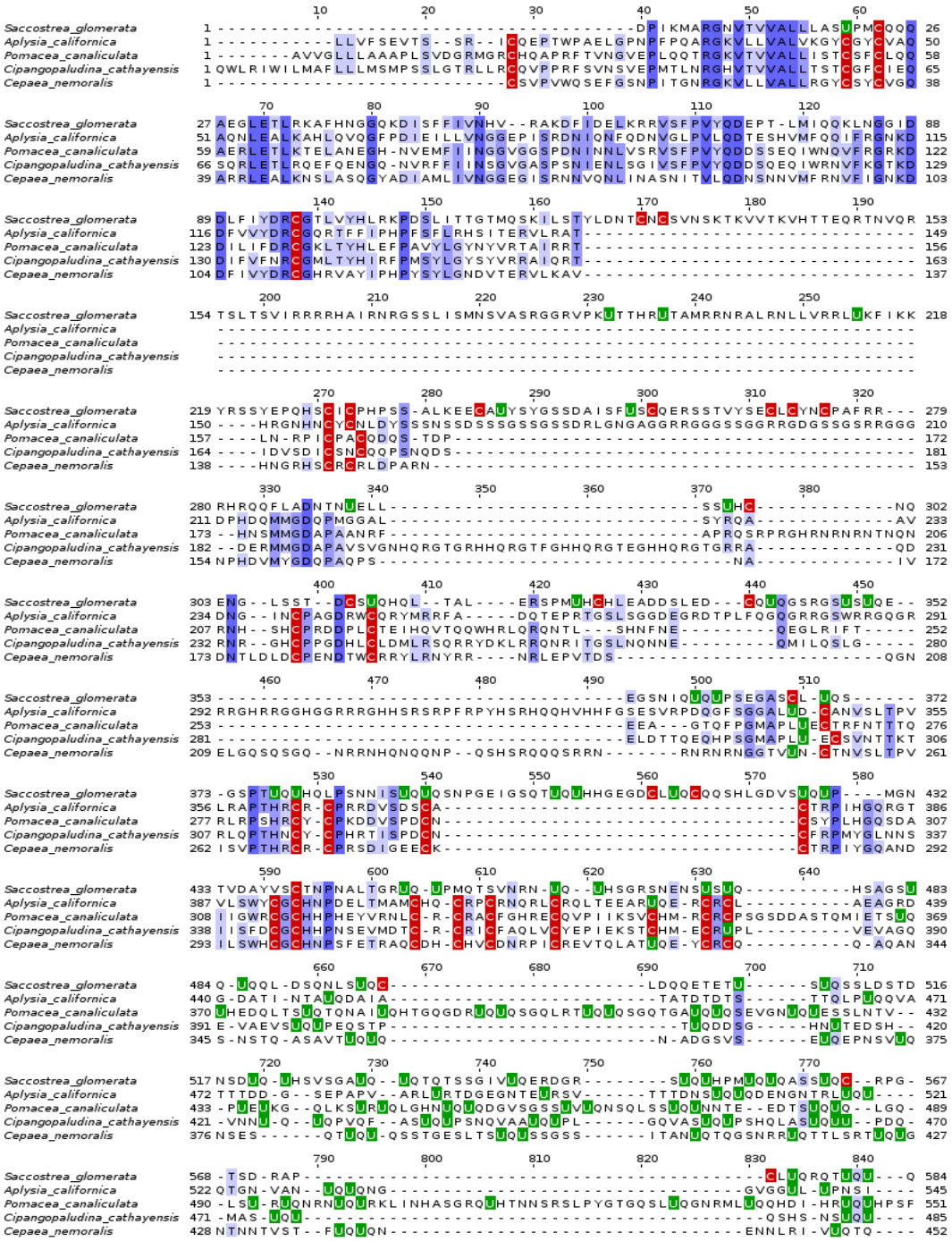


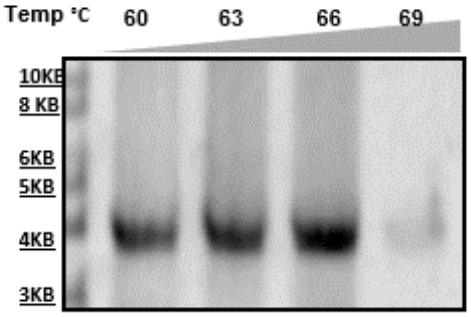
Figure 3.3.2: Gastropod SELENOP protein alignment.

Multiple amino acid sequence alignment containing four SELENOP proteins from representative gastropods with a Cys codon instead of UGA1 (column 59). Cys codons are highlighted in red and Sec codons in green. Blue highlights are conserved amino-acid sequences. Sequences obtained from NCBI TSA: Aplysia californica, GBBW01009782.1; Pomacea canaliculata, GBZZ01023976.1; Cipangopaludina cathayensis, GCEL01087063.1; Cepaea nemoralis, GFLU01113814.1.

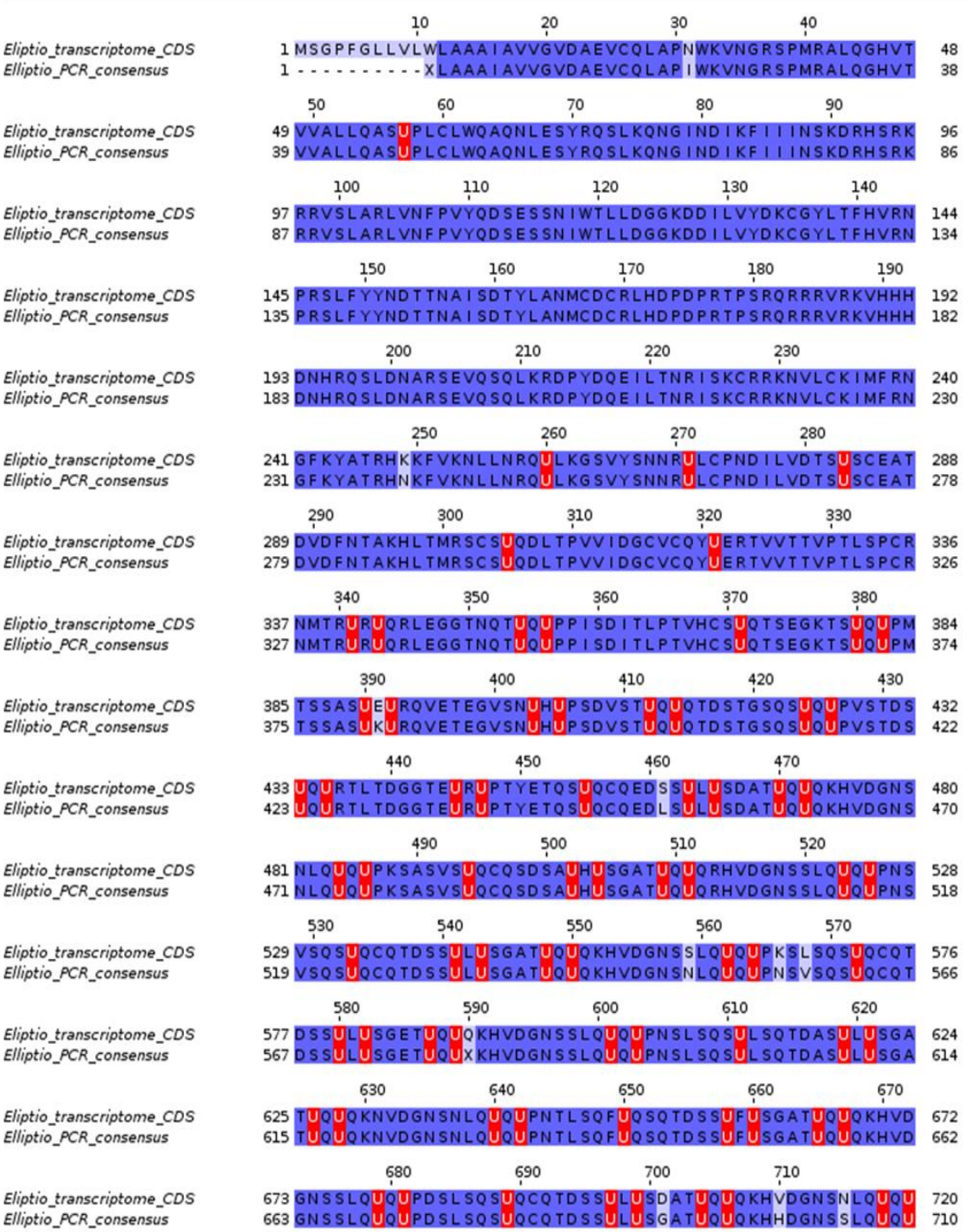
.

Sec-UGA numbers were confirmed from *E.complanata* tissue collected from Maryland USA by RT-PCR and subsequent Sanger sequencing (Fig 3.3.3b). The translated PCR sequence revealed one additional Sec residue matching the position of the Cys residue identified in the published transcriptome (Fig.3.3.3 b, position 738). The original transcript (T1) contained the 5'UTR, the CDS and only SECIS 1 in its 3'UTR (Fig. 3.3.4 a, T1). A further search of SECIS 2 in *E.complanata* using a consensus SECIS 2 sequence from *Margeritefera margeritefera* demonstrated a hit for a shorter transcript in the *E.complanata* transcriptome containing SECIS 2 (Fig. 3.3.4 a, T2). To confirm that *E.complanata* SECIS 1 and SECIS 2 are in a contiguous sequence in SelenoP 3'UTR, an RT-PCR was performed using primers spanning either regions of T1 only for positive control or regions including T2. Multiple products were amplified when a reverse primer against T2 was used indicating non-specific primer binding sites or repeats within the SelenoP transcript sequence. The product detected at >1kb, a product higher than T1 control (Fig. 3.3.4 b) was sequenced. The PCR product that was sequenced using primers binding at different sites within the 3'UTR revealed sequences matching T1 and T2 (Fig. 3.3.4 c) suggesting that they are in a contiguous sequence. In addition, a stretch of TA nucleotide repeats was observed which demonstrates that the absence of SECIS 2 in SelenoP of *E.complanata* was, indeed, an artefact of the transcriptome sequencing assembly.

a)


PCR Gradient Temp °C
60
63
66
69
10KB
8 KB
6KB
5KB
4KB
3KB

b)


Eliptio_transcriptome_CDS
Elliptio_PCR_consensus

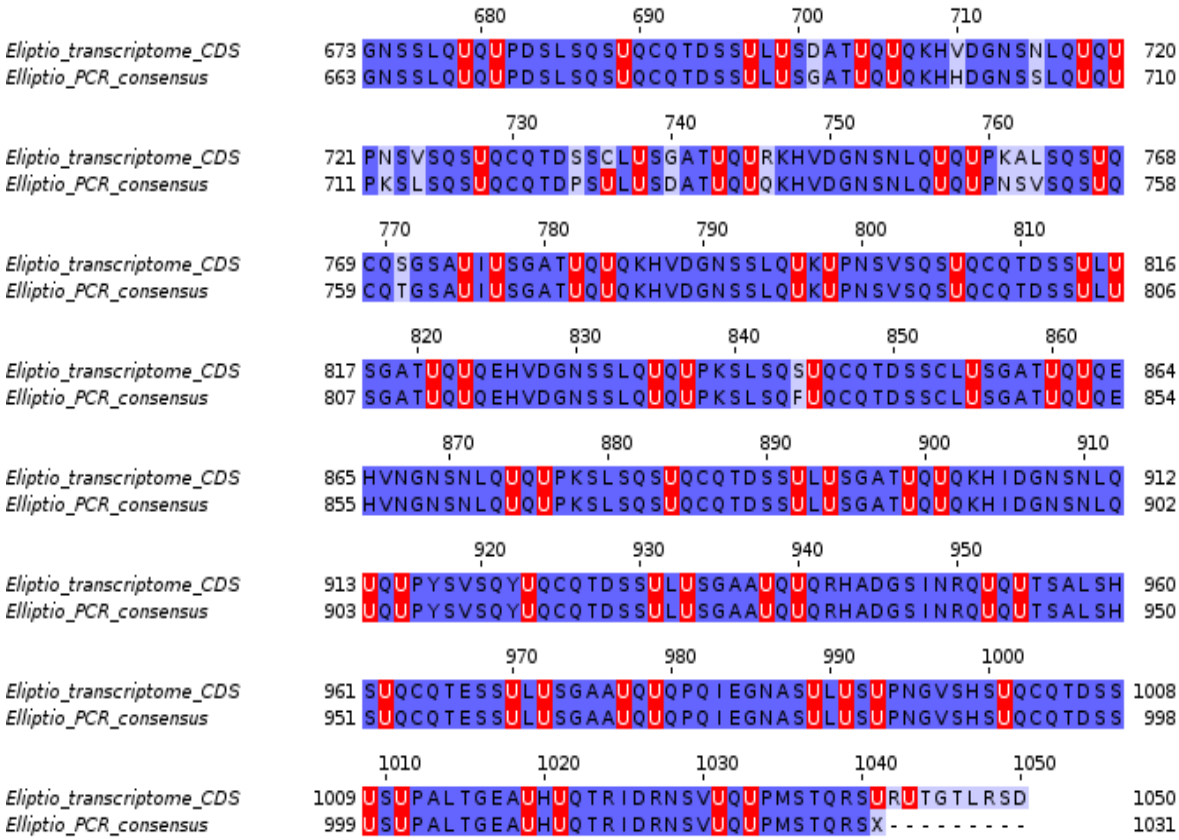


Figure 3.3.3 Confirmation of 132 UGAs in *E.complanata.*

(a) PCR amplification of reverse transcribed cDNA from total RNA of *E.complanata* at varying primer-annealing temperatures. The DNA band was excised, isolated and sequenced by Sanger sequencing. (b) Alignment of PCR sequencing results translated to amino-acid sequences to *E.complanata* protein sequence identified in a published transcriptome. Identified in-frame TGAs (translated as Sec) are in red here depicted as U for selenocysteienes, blue highlights matching residues while light blue are possibly amino-acid substitutions.

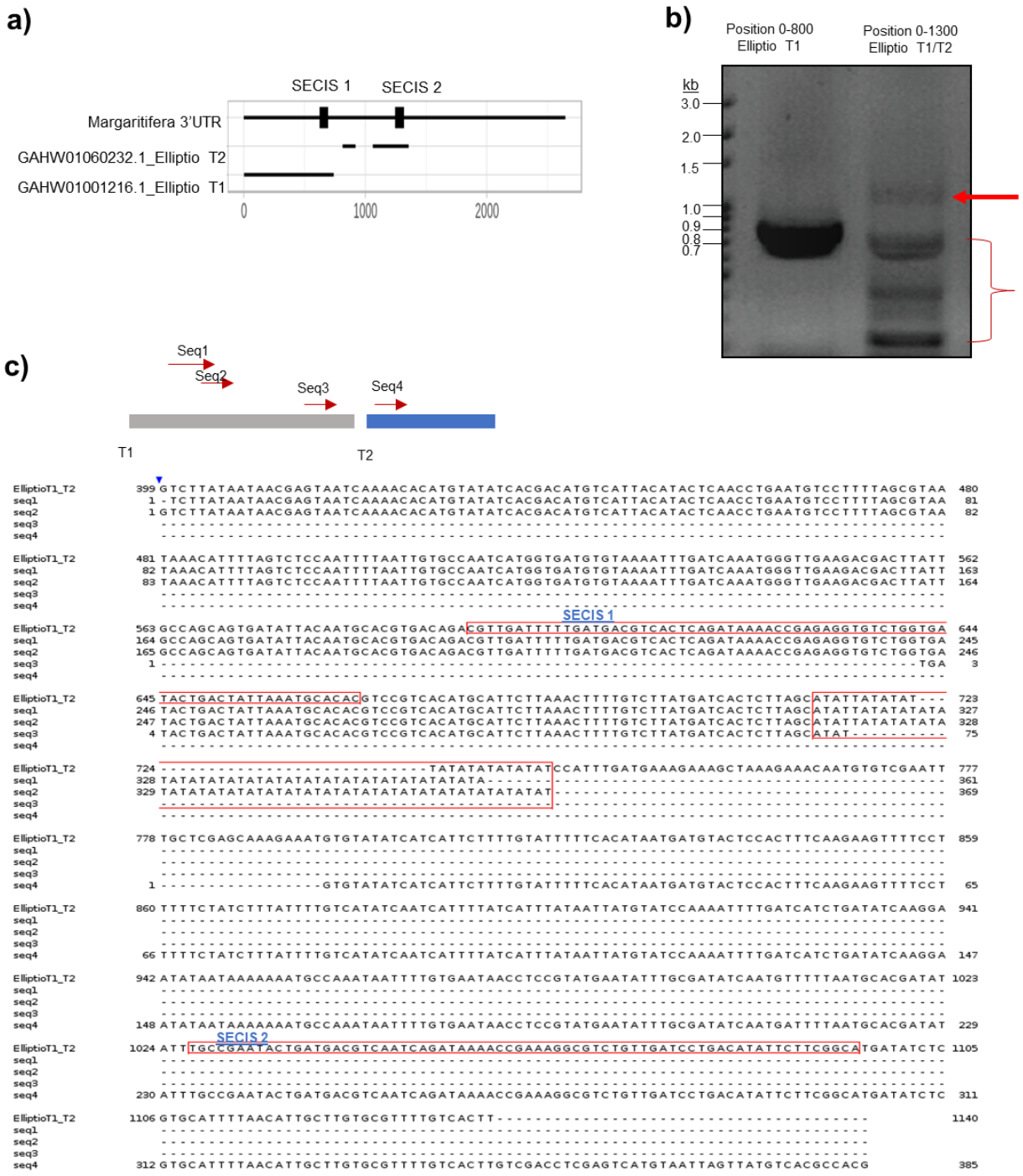


Figure 3.3.4: Confirmation of 2 SECIS elements in *E.complanata*.
(a) *M. margaritefera* 3'UTR alignment with two transcript sequences within the *E.complanata* forward transcriptome. Thick vertical box represent SECIS elements. Eliptio T1 shows a hit containing SECIS 1 while T2 shows a second separate hit containing SECIS 2 from the same *E.complanata* transcriptome. (b) PCR amplification of *E.complanata* SelenoP 3'UTR from cDNA. Lane 1 shows product amplification of primers spanning *E.complanata* T1 from position 0-796 and lane 2 shows product amplification from position 0 to 1300 spanning sequences of T2. Red arrow points to band excised for sequencing while red bracket depicts products from alternative primer binding sites. (c) Nucleotide sequence alignment of sequences obtained from sanger sequencing using S1, S2 and S3 sequencing primers at binding positions indicated on the transcript. Sequences from SECIS 1 and SECIS 2 are highlighted as well as the problematic repetitive TA sequence.

3.3.2: Translation of representative invertebrate SelenoP in a vertebrate reconstituted cell-free translation system

A few representative *selenoP* genes from various metazoan clades were selected for tests of translatability of a T7 polymerase generated mRNA from cloned derivatives. The translation was carried out in rabbit reticulocyte lysate (RRL) coupled with ^{75}Selenium (^{75}Se) labelling (Shetty and Copeland, 2018b) and supplemented with rat CT-SBP2, which is limiting in this system (Mehta et al., 2004, Howard et al., 2007, Donovan et al., 2008, Latreche et al., 2009). The RRL system has been previously shown to support full-length synthesis of vertebrate SELENOP including that of human and zebra-fish with SBP2 supplementation (Shetty et al., 2014) but have not yet been tested to study Sec incorporation of an invertebrate SelenoP.

The mRNA from house spider (*Parasteatoda tepidariorum*), an arachnid arthropod has 9 Sec-UGAs and two SECIS elements (Fig. 3.3.5 a). The owl limpet (*Lottia gigantea*) mRNA, a gastropod aquatic mollusc has 9 Sec UGAs and just one clearly recognizable SECIS element (Fig. 3.3.5 b). The Pacific oyster (*Crassostrea gigas*), a bivalve mollusc, has 46 Sec UGAs and two SECIS elements (Fig 3.3.5 c)

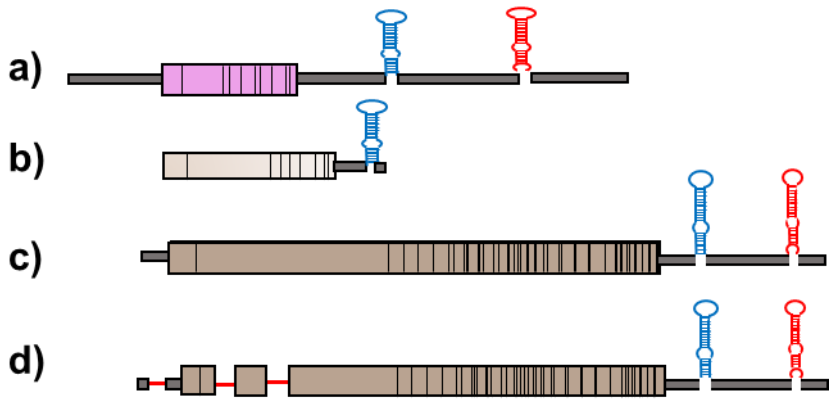


Figure 3.3.5: Native construct designs of representative invertebrate SelenoP.
SelenoP of (a) *Parasteatoda tepidariorum* (house spider), (b) *Lottia gigantea* (sea snail) (c) *Crassostrea gigas* (Pacific oyster) and (d) *C.gigas* genomic DNA sequence to include introns (red line). Coloured boxes are coding sequences, black vertical lines are Sec-UGA, grey thin boxes are UTRs, blue structure is SECIS 1 and red structure is SECIS 2.

Upon translation of the owl limpet mRNA in RRL and subsequent ^{75}Se labelling, no radio-labelled product was observed. The spider mRNA sequence yielded a product detected at around 70 kDa which runs higher than the predicted molecular weight of full-length spider SELENOP (around 50 kDa), possibly due to post-translational modifications. This product is dependent on SBP2 addition, as no radioactive product corresponding to the size of the full-length protein was generated in the absence of SBP2 (Fig. 3.3.6 a). Full-length spider SELENOP also appears to be sensitive to SBP2 concentrations as SBP2 concentrations above 0.64µm shows a significant reduction in radiolabelled product (Fig. 3.3.6 b).

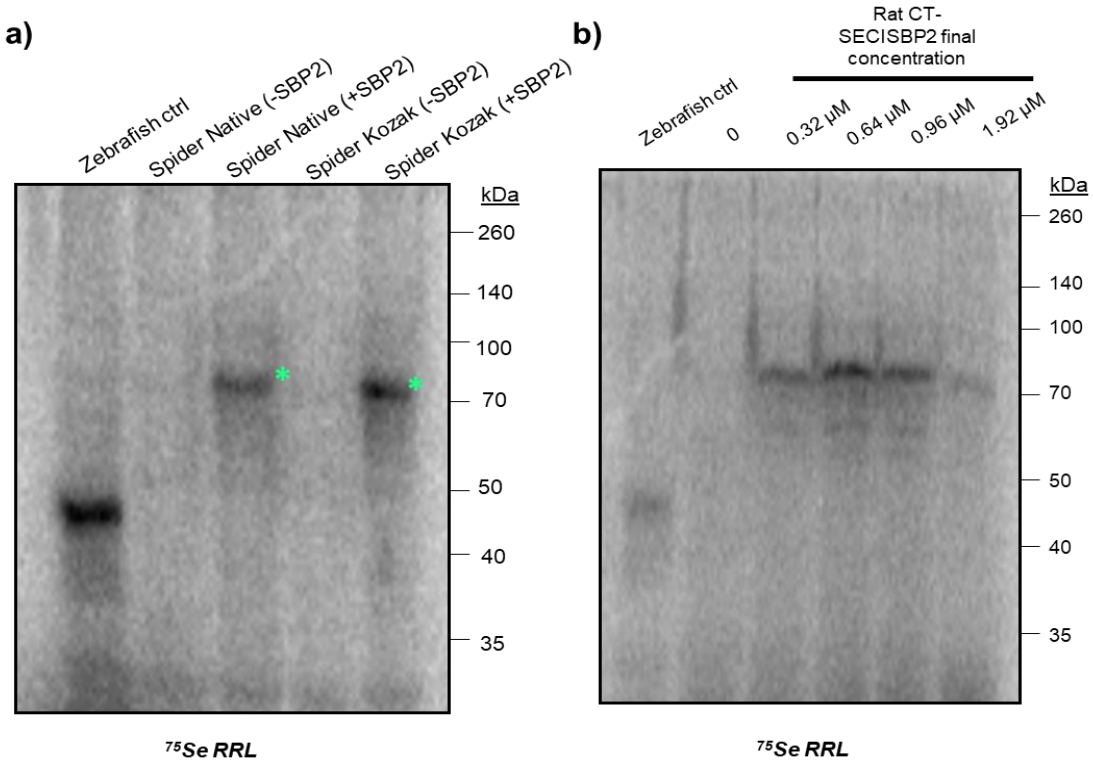


Figure 3.3.6: ^{75}Se labelled RRL translation reactions of spider SelenoP reconstituted with SBP2. (a) RRL translation of native or Kozak consensus-containing spider SelenoP construct with the presence or absence of rat CT-SBP2. Green asterisks denote full-length spider SelenoP. (b) Spider SelenoP RRL translation at varying SBP2 concentrations.

The high Sec content of oyster SelenoP makes it a useful model invertebrate in the study of processive Sec incorporation. A radiolabelled product at an apparent molecular weight of ~30 kDa was synthesized in the reconstituted RRL system which was absent in the negative control without CT-SBP2 (Fig. 3.3.7 a, red asterisk). This corresponds to the molecular weight of SelenoP truncated at approximately its fourth UGA. A CT-SBP2 dependent smearing was also observed (Fig. 3.3.7 a, red bracket). Since UGA positions in the C-terminal of SELENOP is in very close proximity, the smearing could be evidence of premature termination between the UGA-rich domain. Next, oyster SelenoP translation was tested in varying concentrations of CT-SBP2 (Fig 3.3.7 b) and $MgCl_2$ (Fig 3.3.7 c) since SECIS affinity to SBP2 is usually affected by salts and metal ions (Shetty et al., 2018). It was found that no production of full-length SELENOP occurred despite increased concentration of rat CT-SBP2 or $MgCl_2$. Nevertheless, an optimal concentration of $MgCl_2$ (1.5 mM) can improve the amount of translated oyster SelenoP termination product as a stronger intensity of the radiolabelled product at 30 kDa is observed (Fig. 3.3.7 c). CT-SBP2 addition at varying concentrations resulted in the production of another smaller radiolabelled product (around 20 kDa) which decreases as CT-SBP2 concentration is increased (Fig. 3.3.7 b, yellow arrow). This would correspond to the size of oyster SelenoP before it reaches UGA2 and may suggest premature termination or degradation with reconstitution of reduced or non-optimal CT-SBP2 levels.

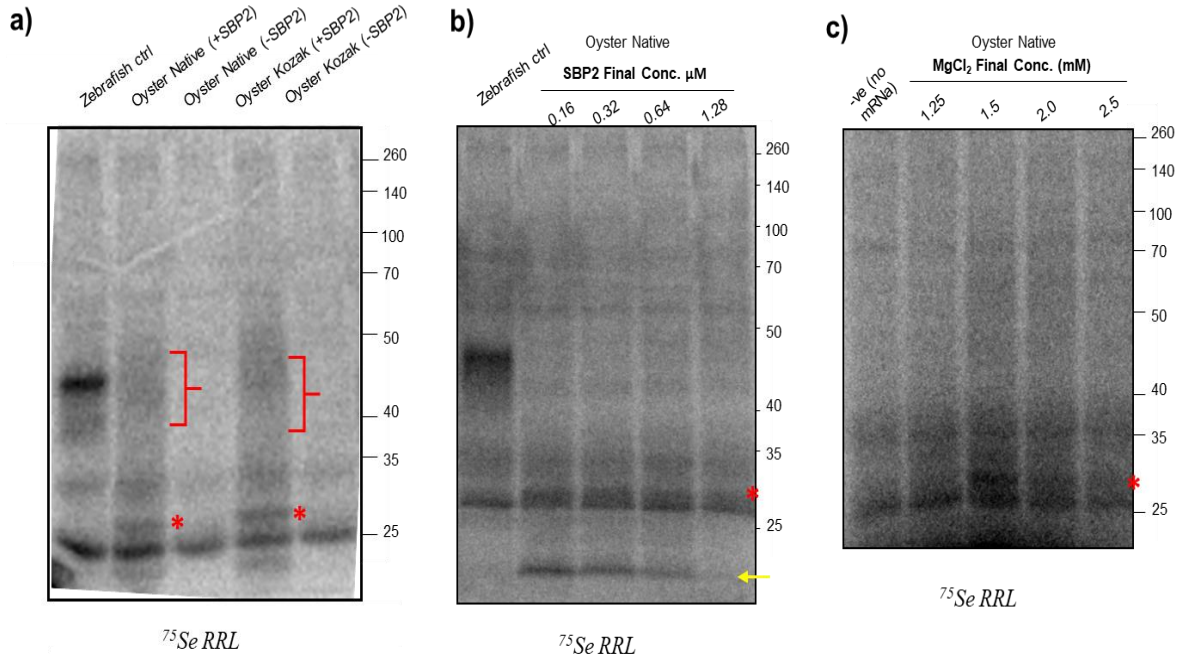


Figure 3.3.7: ^{75}Se labelled RRL translation reactions of Pacific oyster SelenoP reconstituted with SBP2.

(a) RRL translation of native or Kozak consensus-containing oyster SelenoP construct in the presence or absence of rat CT-SBP2. Red asterisks denote oyster SelenoP termination product at UGA 4 and red brackets highlights smearing indicative of termination between UGAs 4-46. (b) Oyster native SelenoP RRL translation at varying SBP2 concentrations. Yellow arrow denotes radiolabelled product indicative of degradation/truncation of the termination product. (c) Oyster native SelenoP RRL translation at varying $MgCl_2$ concentrations. Red asterisks indicate SELENOP termination product observed

Finally, SelenoP translation was tested in cultured HEK293T cells labelled with ^{75}Se (Fig. 3.3.8). Although SELENOP is secreted in the media due to the processing of the N-terminal signal peptide (Hill et al., 1993), both the media and the lysates were tested for SELENOP production. One of the least characterized features of SelenoP is the presence of a conserved intron downstream of UGA1 which may confer potential for nonsense mediated decay. This may also serve as a site for an 'exon-junction complex' recruitment which may have an important role in transcript expression or splicing (Budiman et al., 2011). The translation of an additional oyster SelenoP construct amplified from genomic DNA was tested which includes the intronic sequences of oyster SelenoP (Fig 3.3.5 d) in case an earlier genomic DNA regulation during SELENOP synthesis might occur. Transfection of all SelenoP constructs into cultured cells did not yield any detectable protein product either in the media or in the cell-lysates which could be attributable to either the absence of RNA

expression or protein instability (Fig 3.3.8 a and b). The absence of any detectable SELENOP product in cultured cells suggests a more complex interplay of factors and a more stringent regulation *in-vivo* compared to the RRL system.

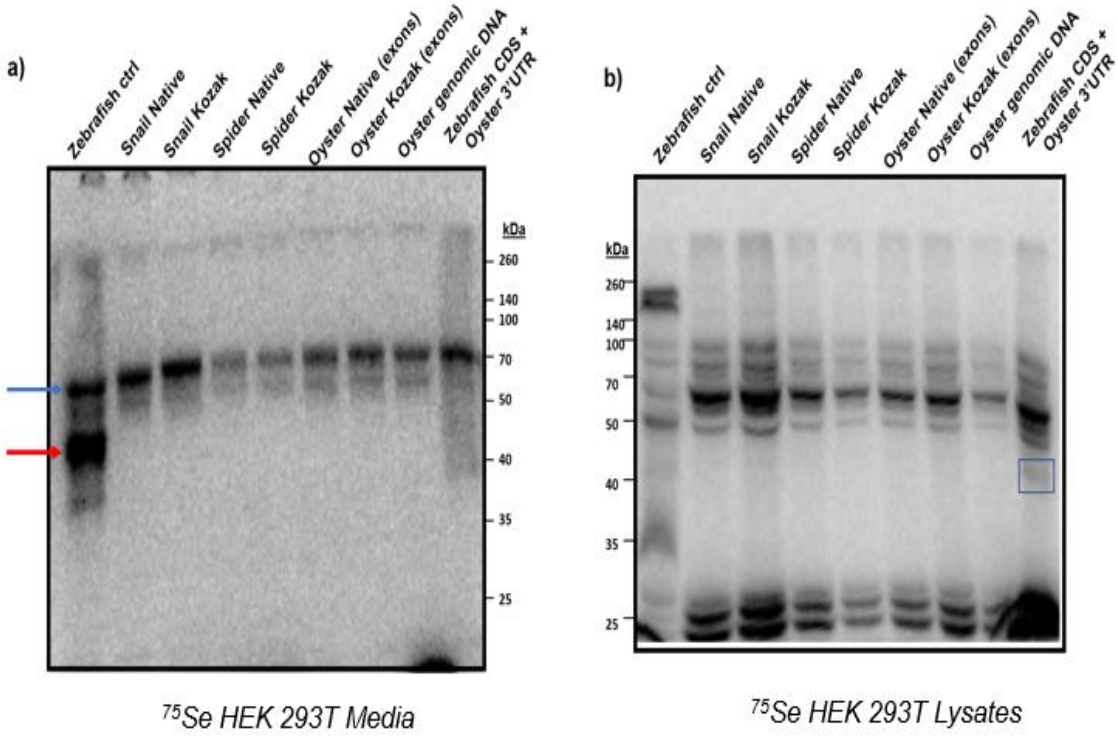


Figure 3.3.8: Autoradiography of invertebrate and mutant SelenoP transfected HEK293T cells. (a) Analysis of SELENOP products from invertebrate and mutant constructs in the cell media 24h after transfection**.** Blue arrow denotes endogenous mammalian SELENOP from the cultured cells while red arrow denotes zebrafish full-length SELENOP. (b) Analysis of SELENOP products from the same samples in the cell-lysates. Blue box denotes a faint zebrafish SELENOP protein upon fusion with zebrafish CDS

3.3.3: Characterization of C.gigas SECIS Binding Protein 2 (SBP2)

A focused experimental analysis was performed on the Pacific oyster, *C.gigas*, SelenoP, because in addition to its important economical role for human consumption, its remarkably high numbers of Sec-UGA is suggestive of a unique decoding feature. Oyster SECIS Binding Protein 2 (SBP2) was characterized since we have previously shown that production of a radioactive termination product from oyster RRL translation is dependent on the presence of rat CT-SBP2.

Sec incorporation is directed by SECIS elements upon recognition of its core motif by SBP2. Prior structure-function analyses have segregated SBP2 protein into the N-terminal domain not required for Sec-incorporation *in-vitro*, a central Sec-incorporation Domain (SID) and a C-terminal RNA binding domain (RBD) which contains the L7Ae RNA binding motif responsible for K-turn motif binding (Fig. 3.3.9 a) (Copeland et al., 2001). Both the SID and RBD are required for Sec-incorporation activity and forms CT-SBP2, while the residues upstream of SID and downstream of RBD are dispensable (Copeland et al., 2001, Budiman et al., 2009). Additionally, C-terminal residues of the SID have been implicated in differential SECIS affinity and mediate stable SECIS interactions necessary for the incorporation of Sec (Bubenik et al., 2013, Donovan et al., 2008). A SBP2 paralogue, termed SBP2-like (SBP2L) was also identified (Copeland et al., 2000) in vertebrates. This also contain putative SID and RBD sites and an additional stretch of a conserved glutamine residues within its C-terminal domain (Fig. 3.3.9 b). A previous survey of eukaryotic SECIS-binding protein (SBP) revealed a sole SBP in invertebrates including sea urchins, sea squirts and the annelid worm *Capitella,* which demonstrates more homology to the SBP2L paralogue in vertebrates (Donovan and Copeland, 2009).

Likewise, in the Pacific oyster, only one copy of SBP2 is present (Fig 3.3.9 c) in contrast to the two copies found in vertebrates (Donovan and Copeland, 2012, Donovan and Copeland, 2009). Alignment of oyster SBP2 against the human SBP2 and SBP2L reveals greater homology of oyster SBP2 to the paralogue SBP2L (Fig. 3.3.9 d), consistent with prior reports in invertebrates. Since reconstitution of rat CT-SBP2 was insufficient to facilitate translation of full-length oyster SelenoP but was able to produce termination products, the effect of substituting oyster SBP2 in the translation of native oyster mRNA in RRL was tested. A recombinant full-length

oyster SBP2 protein was produced and purified from *E. coli* (Fig. 3.3.10). This was supplemented to the RRL translation reaction in place of the rat CT-SBP2.

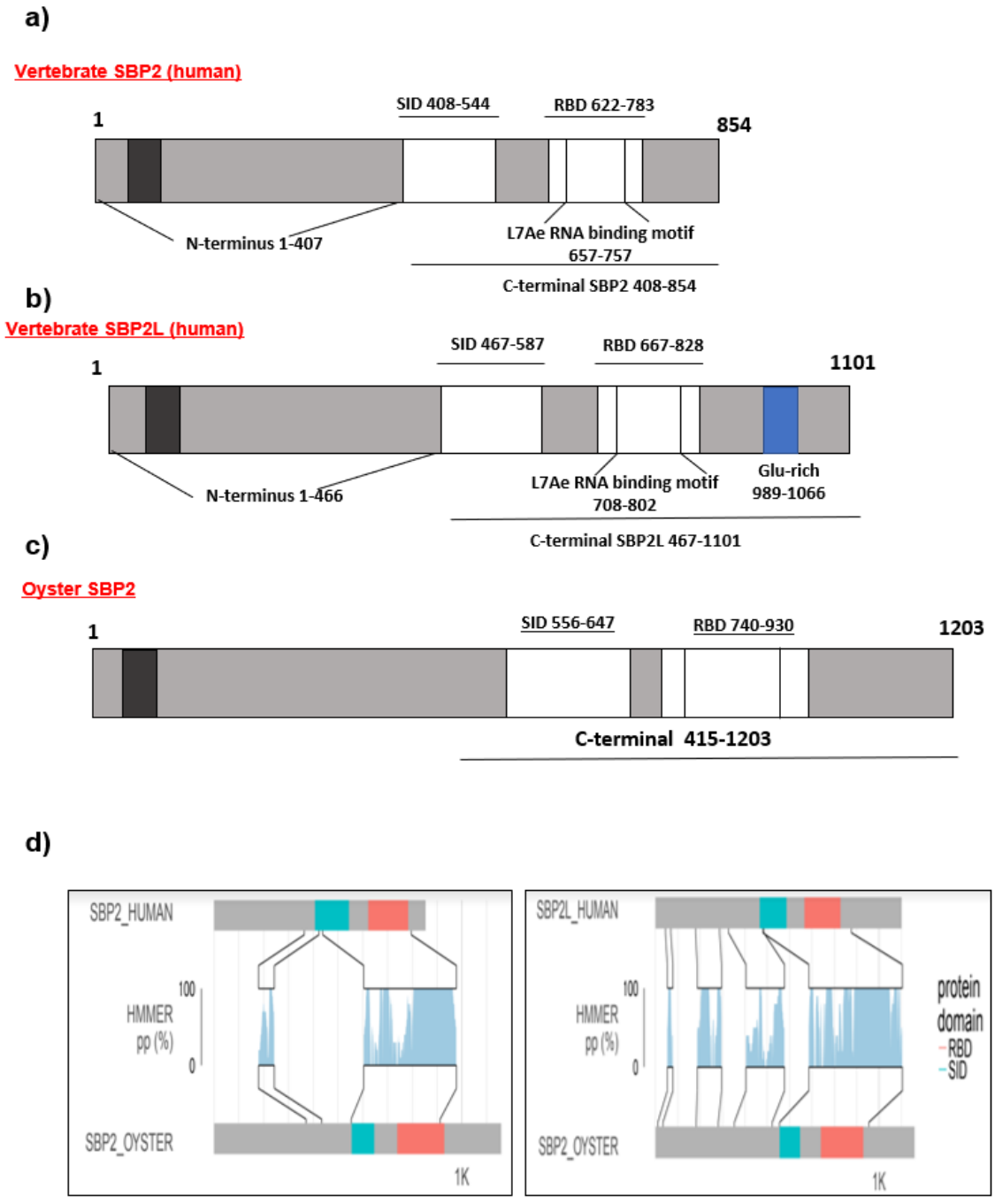


Figure 3.3.9: Comparison of human (vertebrate) SBP2/SBP2L with oyster SBP2.
Protein maps of (a) human (vertebrate) SBP2, (b) human (vertebrate) SBP2L and (c) oyster SBP2. Relative positions of important domains including SID: SECIS Interaction Domain, RBD: Ribosome Binding Domains with the ribosomal protein L7Ae RNA binding motif positions and the glutamine-rich motif (blue) in SBP2L is indicated. The conserved portion of N-terminal domain is in dark grey. (d) Alignment and conservation between human SBP2 or SBP2L and oyster SBP2. (Fig. 3.3.9 a and b are adapted from Donovan and Copeland, 2012 (Donovan and Copeland, 2012)

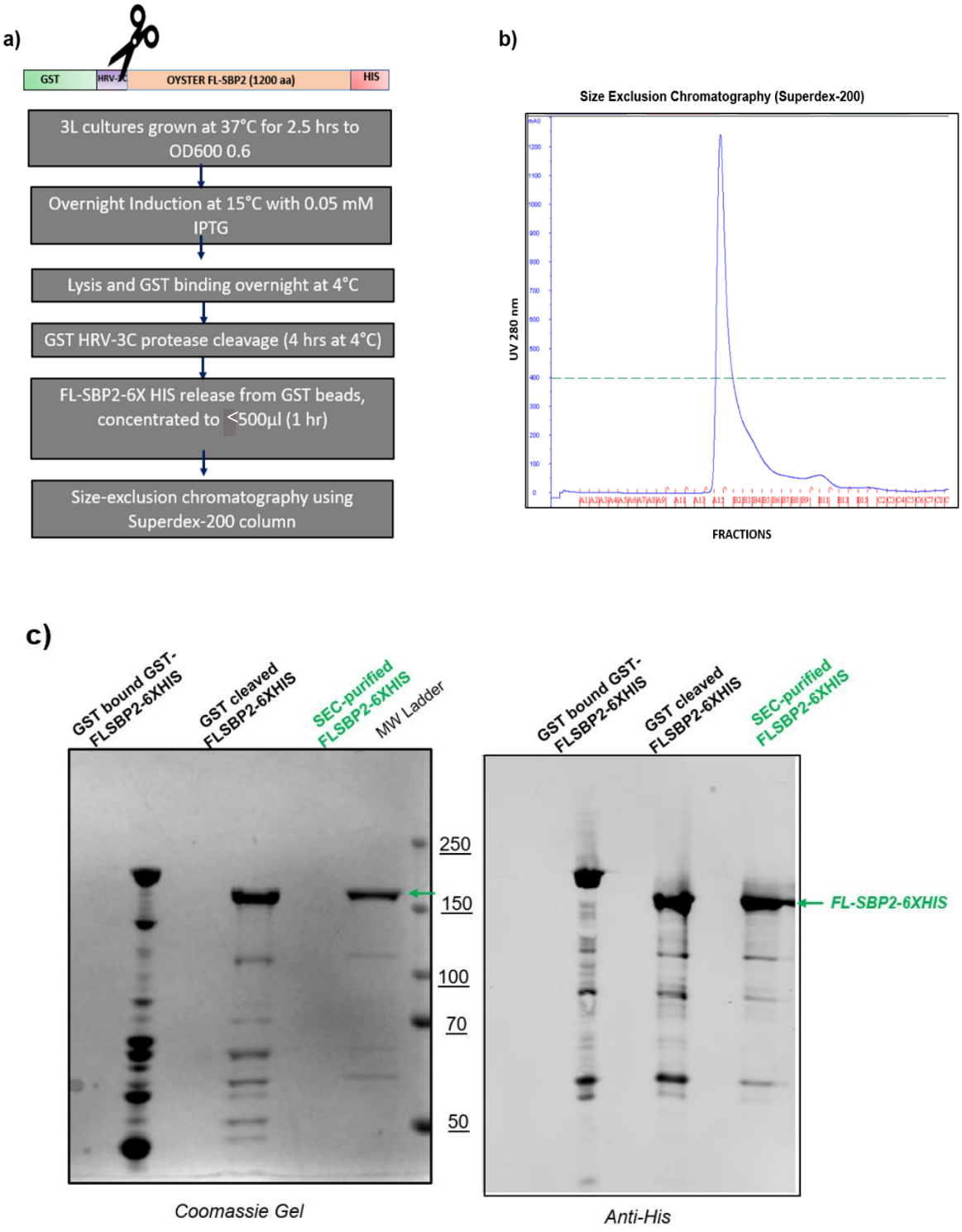


Figure 3.3.10: Recombinant oyster SBP2 purification

(a)Recombinant oyster SBP2 purification work-flow. (b) Size exclusion chromatography peak fraction highlighting the main protein eluting at fractions A15 and B1. (c) Purified oyster SBP2 analysed through SDS-PAGE and subsequent Coomassie staining and anti-His antibody tag probing. The green arrow points to full-length SBP2 with an apparent molecular weight of 170 kDa.

Upon supplementation of full-length oyster SBP2 to the RRL reaction, no protein product was observed for oyster SelenoP (Fig 3.3.11 a) or zebrafish SelenoP (Suppl. Fig 3.5.2 a). In a similar experiment using rat SelenoP mRNA, SBP2-dependent production of a protein product at 27 kDa was observed that would correspond to termination at UGA 2/3 (Fig. 3.3.11 b /Suppl. Fig. 3.5.2 b). Donovan and Copeland previously reported that vertebrate SBP2L was not able to support Sec incorporation while the invertebrate *Capitella telleta* SBP2, with more similarity to the SBP2L paralogue, was able to support a single Sec incorporation for some human selenoproteins (Donovan and Copeland, 2012). This is consistent with the current observation in oyster SBP2, where a single Sec incorporation event can be facilitated but not the processive Sec-incorporation of the UGA-rich distal region.

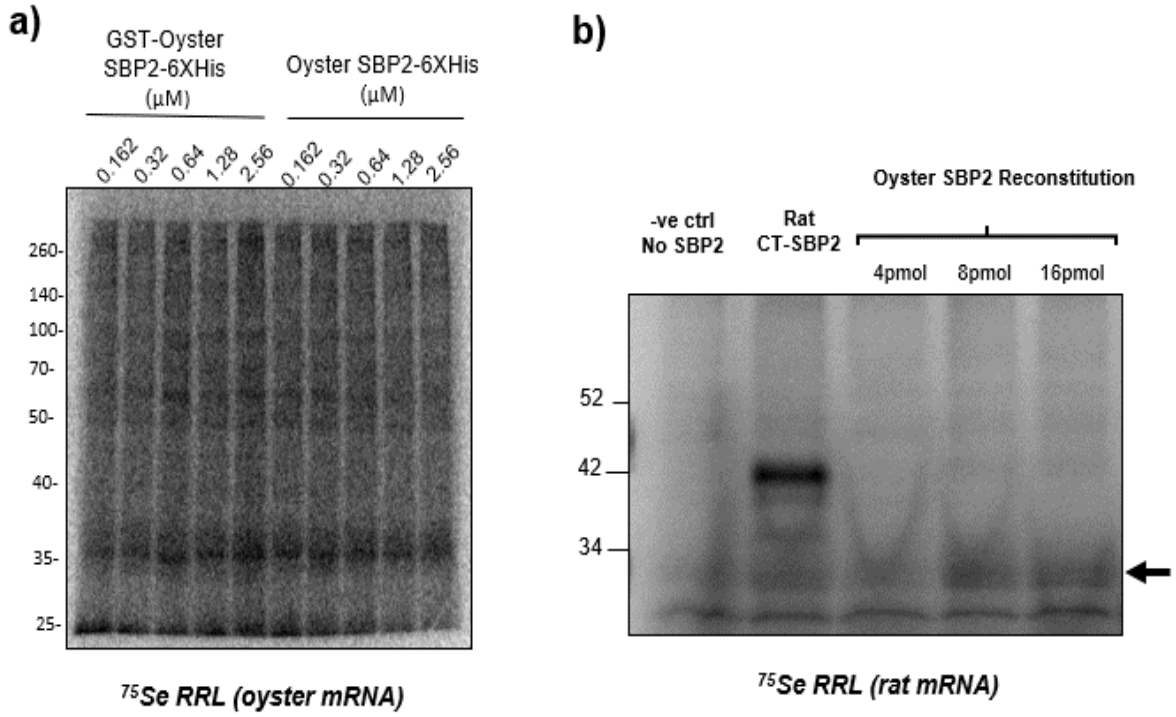


Figure 3.3.11: Reconstitution of purified oyster SBP2 to RRL translation reactions
^{75}Se labelling and RRL translation of (a) oyster SelenoP and (b) rat SelenoP supplemented with varying concentrations of purified recombinant oyster SBP2. The black arrow points to a 27 kDa band suggestive of a single or double Sec-incorporation in rat SelenoP upon oyster SBP2 reconstitution.

3.3.4: Characterization of C.gigas SelenoP mRNA context

Unique features of *C.gigas* SelenoP mRNA was further characterized since it was previously shown that cis-acting signals in mammalian SelenoP are necessary for multi-Sec codon decoding (Mariotti et al., 2017, Shetty and Copeland, 2018a). The mature oyster SelenoP transcript contains two SECIS elements in its 3'UTR with its signature conserved motifs also found in vertebrates (Fig. 3.3.12). Moreover, additional conserved RNA secondary structures along the length of the SelenoP CDS were determined. Among these are the initiation stem loop structure (ISL) which spans positions 37 to 157 from AUG start codon and Sec-recoding elements (SRE 2/3) located between positions 1585-1705 (Fig. 3.3.12). The ISL element is the strongest and most stable structure identified with high conservation in 6 example bivalves (Suppl. Fig. 3.5.3 c); all bivalves containing more than 45 UGAs in their SelenoP.

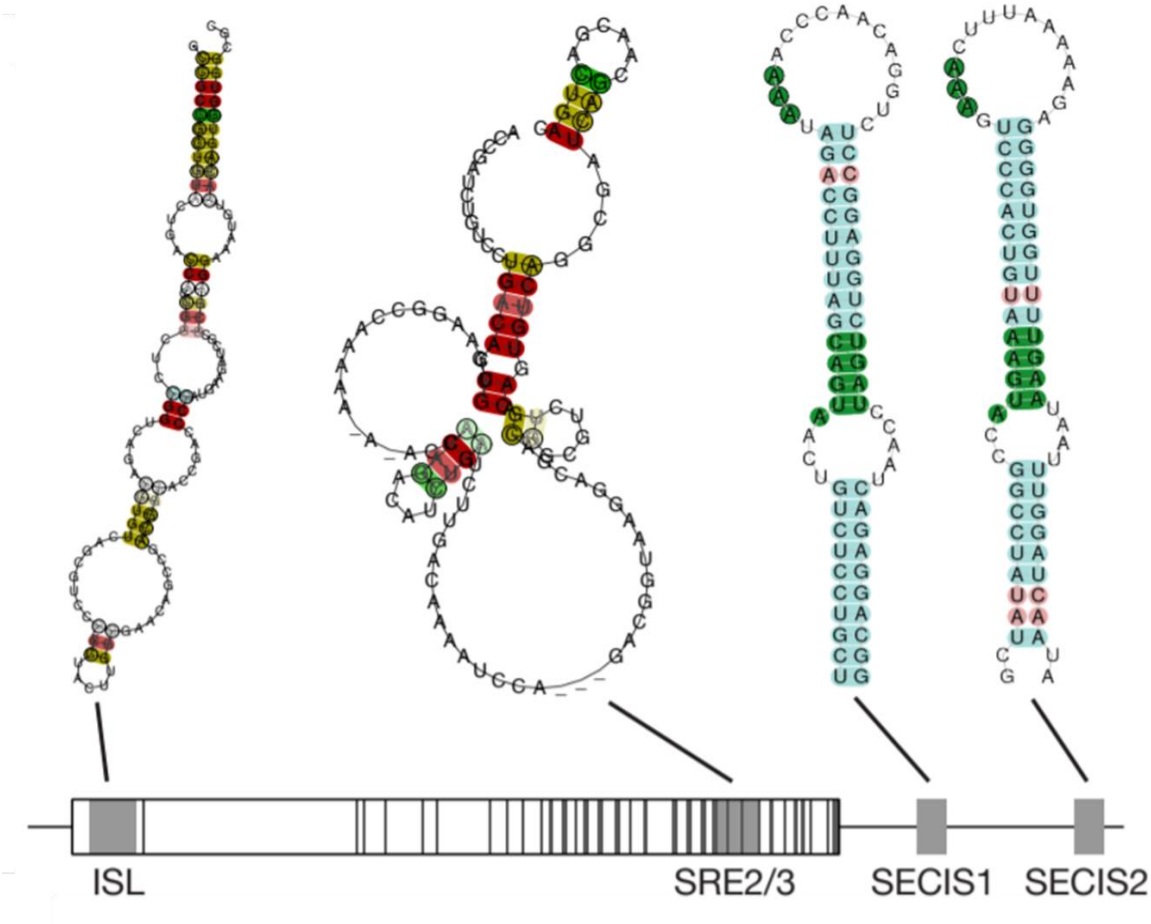


Figure 3.3.12: Conserved RNA structures identified in SelenoP of Pacific oyster *Crassostrea gigas*. SelenoP mRNA indicating the coding sequence, UGA positions (vertical lines), and secondary structure elements (shaded in grey). On top is shown the predicted RNA structure of (left to right) Initiation Stem Loop (ISL) at positions 40-160 from AUG; Selenocysteine Recoding Elements (SRE2/3) spanning positions 1585-1705; and SECIS 1 and 2 at positions 194-267 and 585-657 from UAG stop codon. The conserved motifs of SECIS, namely the core and the A repeats at the apical loop, are highlighted in green.

The effects of context mutation within oyster SelenoP mRNA in the same RRL reconstituted system were tested. Plasmid mutant constructs were generated from the oyster SelenoP WT sequence described above. Capped mutants of SelenoP mRNA containing a T7 promoter were generated from cloned derivatives. First, the poor oyster SelenoP Kozak context c.G(AUG)cG was replaced by a strong Kozak context AcC(AUG)Gca. There was no qualitative product difference, but a modest increase in the amount of the termination product (Fig 3.3.13, O2 and Fig 3.3.7a). On removal of the 5'UTR (Fig 3.3.13, O3) there was a modest reduction of the amount of the termination product. Synonymous mutation of codons of the ISL to disrupt

complementarity (Fig 3.3.13, O4) (disrupted structure represented in Suppl. Fig. 3.5.3b) led to absence of detectable ^{75}Se product. Although this does not consider termination product at UGA1 as SelenoP translation would result to a 5 kDa, non-labelled product beyond detection of the described assay, the abolishment of the termination product upon ISL mutation suggest the importance of the ISL element in SelenoP translation (either at initiation or for UGA 1 decoding to Sec). Removal of the sequence from UGA11 to UGA46 yields no SelenoP-specific radiolabeled product (Fig. 3.3.13 O5). While this deletion removes the Sec redefinition elements (SRE 2/3) spanning UGAs 31-35, the lack of labelled product likely reflects long range mRNA folding being relevant to selenocysteine specification.

The oyster SECIS elements exhibit the same conserved K-turn motif as in their vertebrates counterparts (Walczak et al., 1996) consisted of GA:GA base pair (Fig 3.3.13, green) which is recognized by SBP2 (Copeland et al., 2000, Copeland and Driscoll, 1999, Lesoon et al., 1997). The conserved stretch of adenines is also found in the apical loop. The oyster SECIS elements show functionality with substantial affinity to rat CT-SBP2 as it was able to facilitate translation of zebrafish SelenoP CDS upon fusion to the full oyster SelenoP 3'UTR. In cultured cells, however, fusion of oyster 3'UTR to zebrafish SelenoP CDS affected protein export resulting to the absence of SelenoP product in the media (Fig 3.3.8 a, final right lane). Zebrafish protein product was found in reduced abundance in the lysate fraction (Fig 3.3.8 b, blue box), again highlighting the more stringent control in cells. Strikingly, it was observed that fusion of individual oyster SECIS 1 or oyster SECIS 2 to the zebrafish SelenoP coding region, translated in RRL, yielded full-length zebrafish SELENOP product to the same efficiency (Fig 3.3.13, Lane O7, O8, O9).

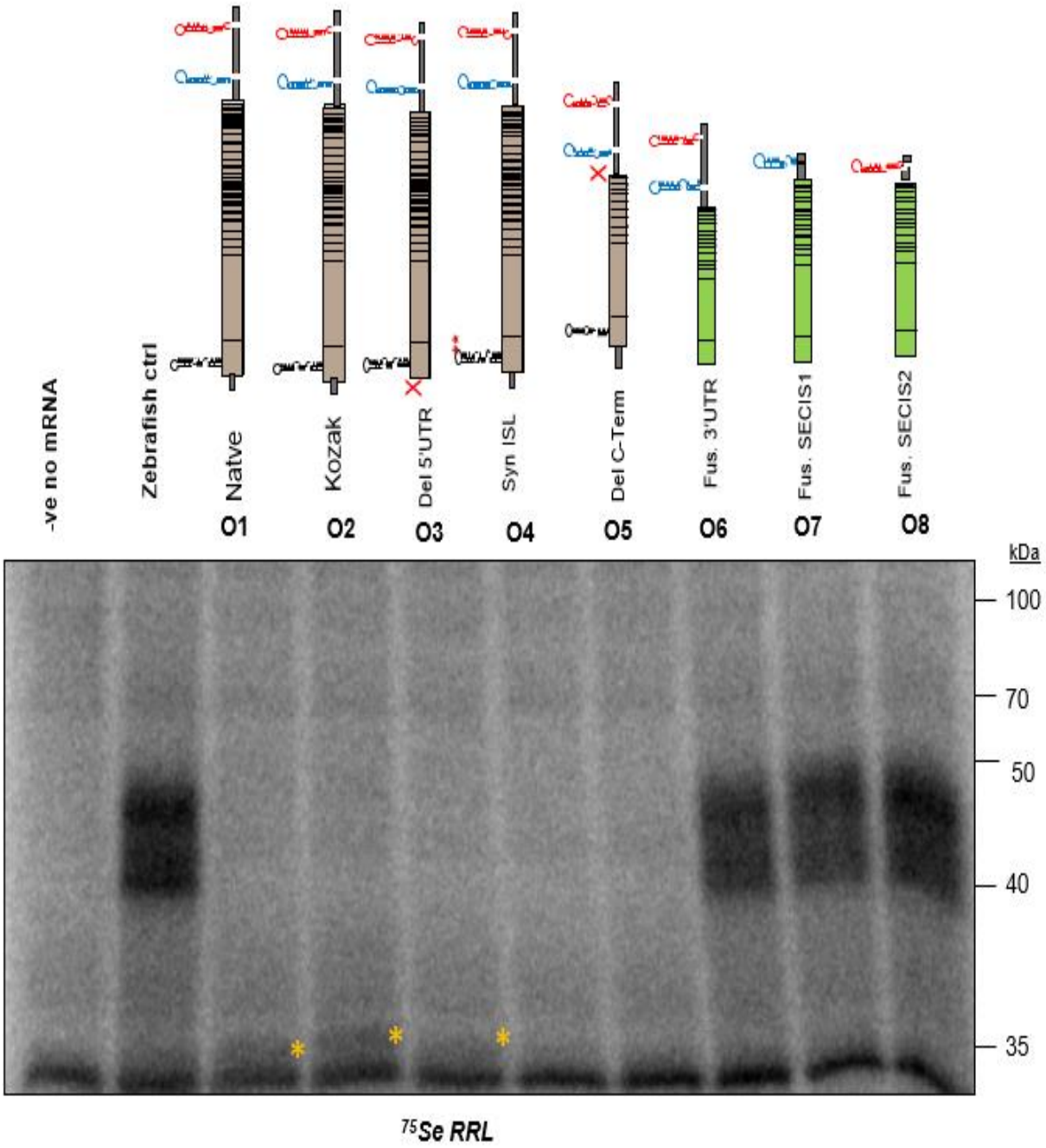


Figure 3.3.13: Oyster SelenoP mRNA context-specific mutants for RRL translation.
Brown box is oyster CDS and green box is zebrafish CDS. Black lines are Sec-UGA, structure before UGA 1 is ISL structure, SECIS 1 is in blue and SECIS 2 is in red. Red 'X' denotes deletion and '**' denotes synonymous codon substitution. Changes in the oyster SelenoP include: O1: native, O2: Kozak addition, O3: 5'UTR removal, O4: ISL disruption by synonymous codon mutation, O5: removal of UGA 11-46. O6-O8 are fusion to zebrafish CDS: O6: fusion to oyster 3'UTR, O7: fusion to oyster SECIS 1, O8: fusion to SECIS 2. The yellow asterisk denotes oyster SELENOP termination product.

Prior reports have suggested that RNA kink-turn motifs are polymorphic and metal-ion dependent. It can exist in an equilibrium between a strongly-kinked conformation or a weaker bulge-like structure (Goody et al., 2004). We tested oyster SECIS fusion constructs for their ability to drive processive Sec-incorporation (possibly by improved interaction with SBP2) with increasing $MgCl_2$ concentration. Fused oyster individual SECIS elements or full 3'UTR to zebrafish SelenoP CDS (Fig. 3.3.14 b, c and d) shows similar efficiency to produce full-length SELENOP, in comparison to the wildtype version of zebra-fish SelenoP containing its native 3'UTR (Fig. 3.3.14 a) despite the addition of increased $MgCl_2$. Our findings from the oyster SECIS fusion constructs differs from the established findings in vertebrate of a differential role for the two SECIS elements (Berry et al., 1993, Stoytcheva et al., 2006)**.** Moreover, recent reports studying SECIS functionality in the same *in-vitro* system have shown that for SelenoP mRNA, SECIS 1 alone can support full processive Sec incorporation while SECIS 2 can only incorporate UGA 1 *in-vitro* (Shetty et al., 2018), similar to results identified in mice (Wu et al., 2016). This is contrary to the lack of functional differentiation demonstrated by the oyster SECIS elements in the heterologous system it was tested.

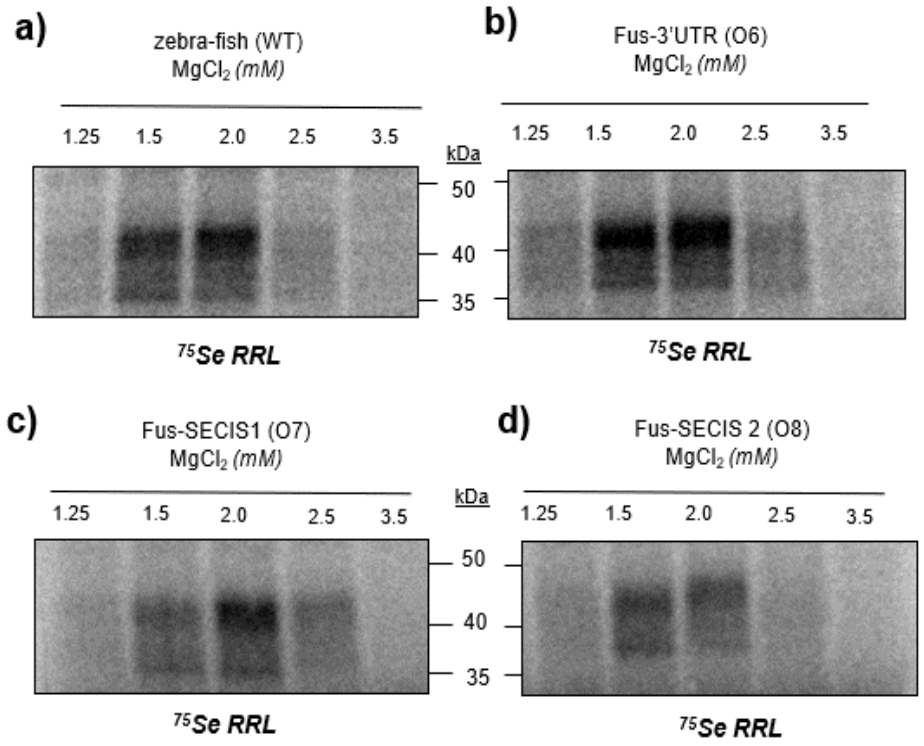


Figure 3.3.14: $MgCl_2$ assay of oyster SECIS fusion constructs' affinity to rat CT- SBP2 for processive incorporation.

Addition of increasing concentrations of $MgCl_2$ (mM) to RRL reactions of (a) wild-type zebrafish (b) O6: zebrafish fusion with oyster 3'UTR (c) O7: zebrafish fusion with SECIS 1 and (d) O8: zebrafish fusion with SECIS 2.

3.4 Discussion

SelenoP evolution across metazoans

SelenoP is a fundamental factor in selenium homeostasis (Burk and Hill, 2015) as its levels are linked to selenium supply. At the level of individuals, SelenoP levels in plasma is a biomarker of selenium concentration (Xia et al., 2005, Kipp et al., 2015) while at the species-level, the number of Sec in SelenoP correlates to selenoproteome size and with the degree of Se utilization (Lobanov et al., 2007).

The evolutionary analysis of SelenoP presented revealed remarkable diversity across metazoans. The N-terminal thioredoxin-like domain containing the UxxC motif in SelenoP (U for Selenocysteine) was found conserved across species apart from gastropods in which the Sec was replaced with Cys. In addition to this, only a single SECIS element was found in gastropod SelenoP supporting the 3'UTR looping model placing SECIS 2 in close proximity to UGA 1 and facilitating decoding to Sec during SelenoP translation (Stoytcheva et al., 2006).

The C-terminal domain containing multiple Sec-residues demonstrate a high degree of variability often posing problems in the phylogenetic construction due to issues in motif-recognition. This multi-Sec portion of SELENOP demonstrates modularity in several invertebrate lineages. Certain motifs are repeated, and their occurrences are more similar to other occurrences within the same gene than to the sequence of SelenoP in other metazoan groups. Such observations indicate that independent events of elongation occurred to extend the C-terminal tail by repetition of the basic Sec-containing module. In other lineages, this Sec-rich domain was either shortened or lost entirely (e.g. insects). Cumulatively, these analyses suggest a dynamic process acting on the C-terminal tail of SELENOP characterized by fast divergence, extension or contraction. Such process was probably driven by the environmental availability of selenium across biological niches, as well as the changing reliance on selenium utilization by the various metazoan lineages.

From previous vertebrate analysis, it was found that aquatic organisms possess a larger selenoproteome and more Sec in their SelenoP compared to terrestrials (Lobanov et al., 2008). Such observations were linked to increased Se-requirement in aquatic environments and a reduced Se-reliance in terrestrial habitats (Lobanov et al.,

2007). In addition, recent reports show that terrestrial vertebrates have a more relaxed evolutionary constraints in their selenoprotein genes compared to teleost fish (Sarangi et al., 2018) as the C-terminal portion of terrestrial vertebrates SELENOP appear to follow an almost neutral process of Sec to Cys conversion. Analyses of metazoan SelenoP are very much consistent with the increased utilization of Se in aquatic species hypothesis. While there are some exceptions (e.g. tunicates and crustaceans lacking SelenoP), the Sec-richest proteins were all found in aquatic species. Certain aquatic lineages demonstrating this exception may have evolved a protein of a similar function. Another explanation could be that their putative SELENOP may have diverged excessively beyond homology-recognition from the ancestral SELENOP form. Nevertheless, the results show that SelenoP evolution led to truly outstanding number of Sec residues in marine worms and molluscs, with freshwater mussel *E. complanata* topping the list, containing 131/132 Sec-UGAs.

The translation components for oyster SelenoP Sec-incorporation

Cell-free translation systems, in particular the rabbit reticulocyte lysate and the wheat germ lysate systems, have been invaluable tools for researchers to determine the factors necessary for Sec-incorporation. Studies of Sec incorporation in selenoproteins have expanded significantly in RRL (Copeland et al., 2000) since the observation that SBP2 is extremely limiting in this system. SelenoP studies of Sec-incorporation in the RRL were particularly insightful, however, they have always focused on vertebrate SelenoP characterization.

Since little is known on SelenoP outside vertebrates, we experimentally tested SelenoP translation of representative invertebrate species identified from these analyses (arthropod arachnid, mollusc bivalve, mollusc gastropod) in RRL. We found that *in-vitro* RRL reconstituted with mammalian components of Sec-machinery can facilitate Sec incorporation in some invertebrate SelenoP. For instance, full-length translation of spider SelenoP with 9 Sec residues is facilitated by rat CT-SBP2 reconstitution in RRL suggesting that all the factors necessary for spider selenoP translation are present in this system. The attempts to translate mollusc SelenoP in reconstituted RRL were less successful. No product was observed for owl limpet, perhaps linked to its apparent absence (or increased divergence) of a second SECIS in its 3'UTR.

Introduction of the oyster mRNA in RRL yielded only a termination product that corresponds to termination at the earlier distal UGAs (3 or 4) indicating that processive Sec incorporation is not supported here. A strong and stable RNA secondary structure corresponding to the initiation stem loop also identified in mammals (Mariotti et al, 2017) was found in oysters and was conserved across bivalves (Fig 3.3.12 a). Its disruption led to the absence of any ^{75}Se radiolabelled product. As the RRL assay is unable to detect termination at UGA1, we are unable to say whether ISL disruption abrogated initiation or the ability to decode UGA as selenocysteine. Moreover, deletion of a substantial part of the C-terminal domain (UGA 11-46) and retainment of initiation context did not yield SelenoP product, suggesting alternative regulatory elements or long-range interactions necessary for Sec-redefinition. More recently, studies have suggested important coding region determinants of SelenoP in human and zebrafish which are required for efficient Sec incorporation (Shetty and Copeland, 2018c). SelenoP transfection in cultured cells for all the native constructs tested did not yield any detectable protein product suggesting a more stringent control of SelenoP synthesis in cells either at the level of transcript expression or translation.

Perhaps, the oyster SECIS elements had the most surprising effect on Sec-incorporation as fusion of either SECIS 1 or SECIS 2 to zebrafish SelenoP CDS was able to facilitate full-length translation. This is contrary to the wild type SECIS function of zebrafish where SECIS 1 is sufficient for full-length SelenoP translation while SECIS 2 is only able to incorporate a single Sec (Shetty et al., 2018) consistent with the model of SECIS differential regulation of UGA decoding in SelenoP (Stoytcheva et al., 2006). With caution in extrapolating from this heterologous system, this may mean that oyster SECIS 1 and 2 are more functionally interchangeable than their mammalian counterparts (Wu et al., 2016).

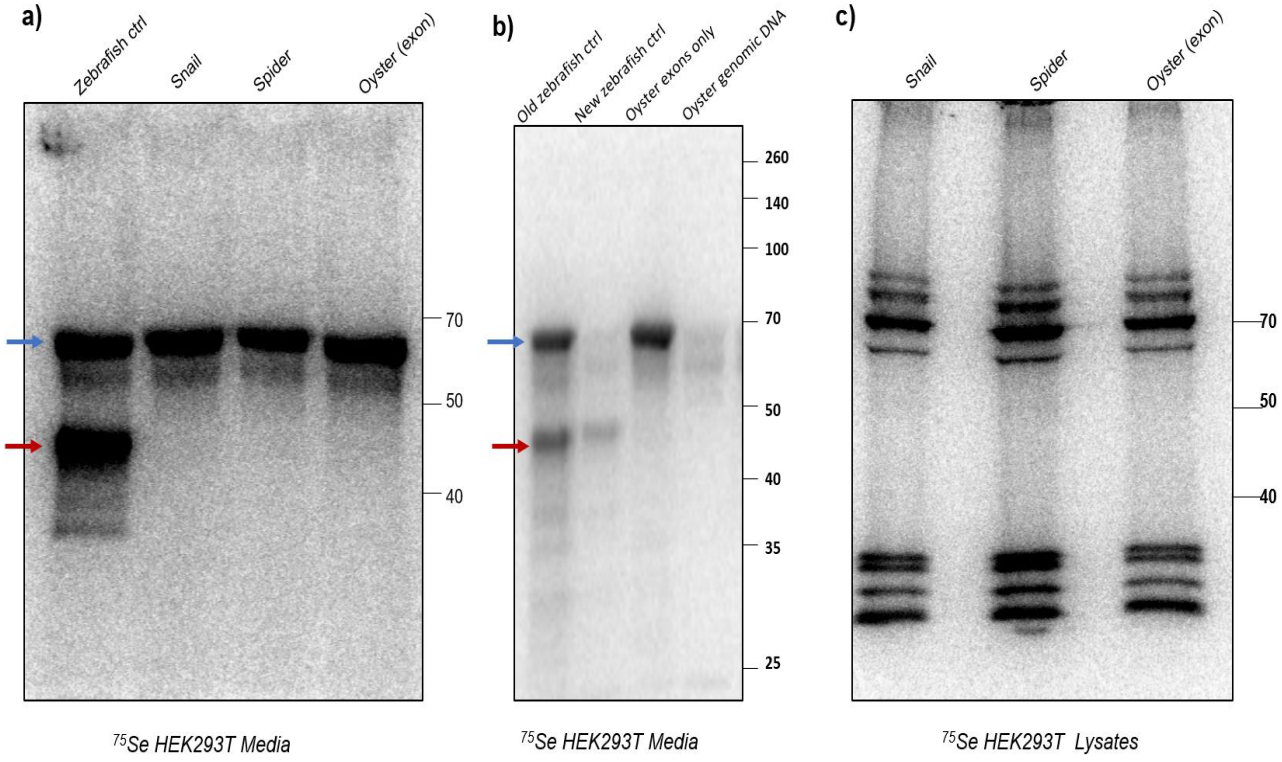
Moreover, characterization of oyster SBP2 revealed higher homology to the vertebrate SBP2L paralogue, thus, it was not surprising that *in-vitro*, oyster SBP2 was unable to substitute for rat CT-SBP2 in promoting specification of multiple Sec. Such observations supported previous findings (Donovan and Copeland, 2012). Ribosome profiling of mice with conditional deletions of SBP2 demonstrated that SBP2 is not strictly required for Sec incorporation *in-vivo* as translation and Sec-redefinition of selenoproteins remained unaffected despite significant reduction in mRNA-levels

(Fradejas-Villar et al., 2017) suggestive of an alternative function in mRNA stability. Oyster SBP2/SECIS interaction might not be optimal *in-vitro*. However, it is unclear why oyster SBP2 has better affinity to rat SelenoP (SECIS elements) and can support a single Sec incorporation event but not with oyster or zebrafish mRNA (Suppl. Fig. 3.5.4).

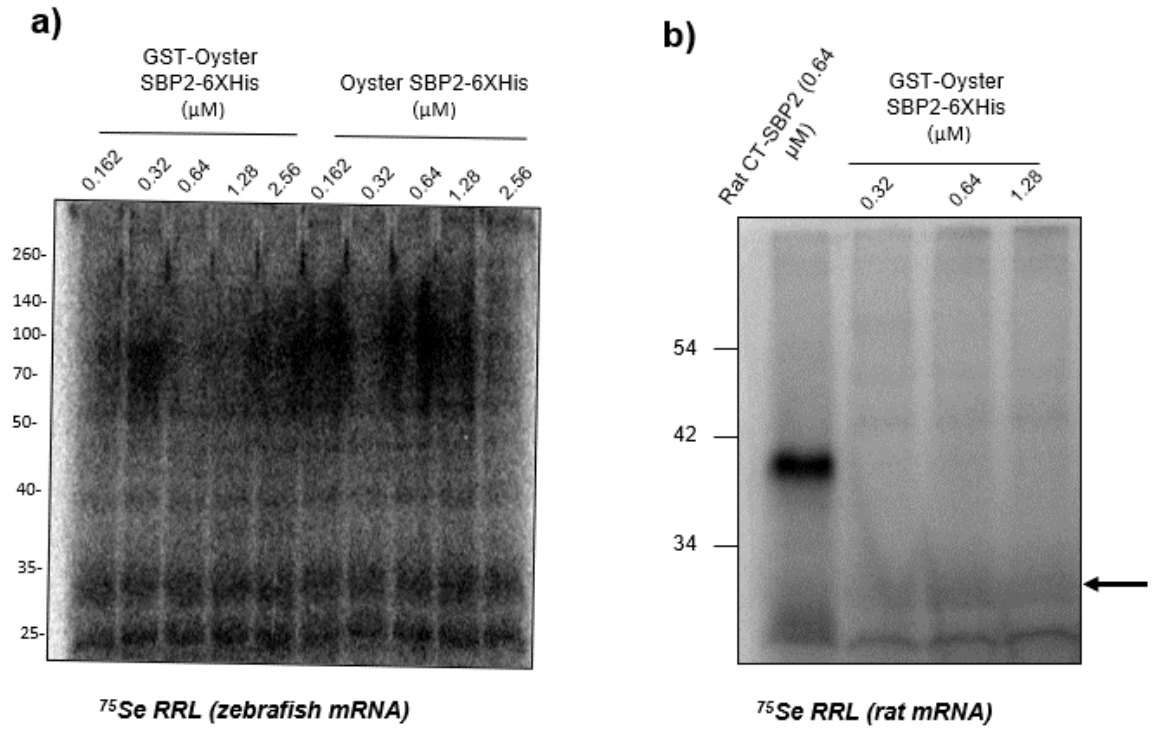
The understanding of essential Sec incorporation factors along with SelenoP mRNA determinants provided better insights into the complexity of SelenoP regulation. Our results here highlight a degree of interchangeability between Sec incorporation factors and SelenoP mRNA elements across invertebrates, mammals and fish. For instance, the spider SelenoP can be translated using mammalian Sec-machinery components and the oyster SECIS elements can be used to drive processive incorporation in zebrafish SelenoP CDS. Synthesis of such termination products from the oyster SelenoP and incompatible SBP2 interaction, indicates insensitivity of the vertebrate translation components to signals within oyster SelenoP mRNA and/or unknown oyster trans-acting components lacking in the RRL, necessary for processive Sec incorporation. While studies using RRL reconstitution were informative, some fundamental question remained unresolved in this chapter: is the full-length oyster SelenoP translated *in-vivo* and does Sec-incorporation occur in response to UGA in the oyster? What role does Se play in this process? We thus performed homologous *in-vivo* studies in oysters and discussed further insights gained in Chapter 4.

In addition to RRL, the wheat germ lysate system (derived from wheat germ plants that are void of any Sec-incorporation factors) has tremendous potential to answer fundamental mechanistic questions. From this system it has been established that SBP2, eEFSec and Sec-tRNA^{SerSec} were sufficient to promote Sec-incorporation and that mammalian ribosomes were not required (Shetty et al., 2014). It was also recently found that while RRL can support full-length Sec incorporation in tested vertebrate SelenoP, only a single Sec-incorporation event was observed in wheat germ lysate (Shetty et al., 2014) highlighting an unidentified processivity factor. In the case of the oyster, the wheat germ lysate might be an important tool to understand what drives ribosome specialization in such a unique multi-Sec decoding process.

3.5 Supplementary Figures/Tables

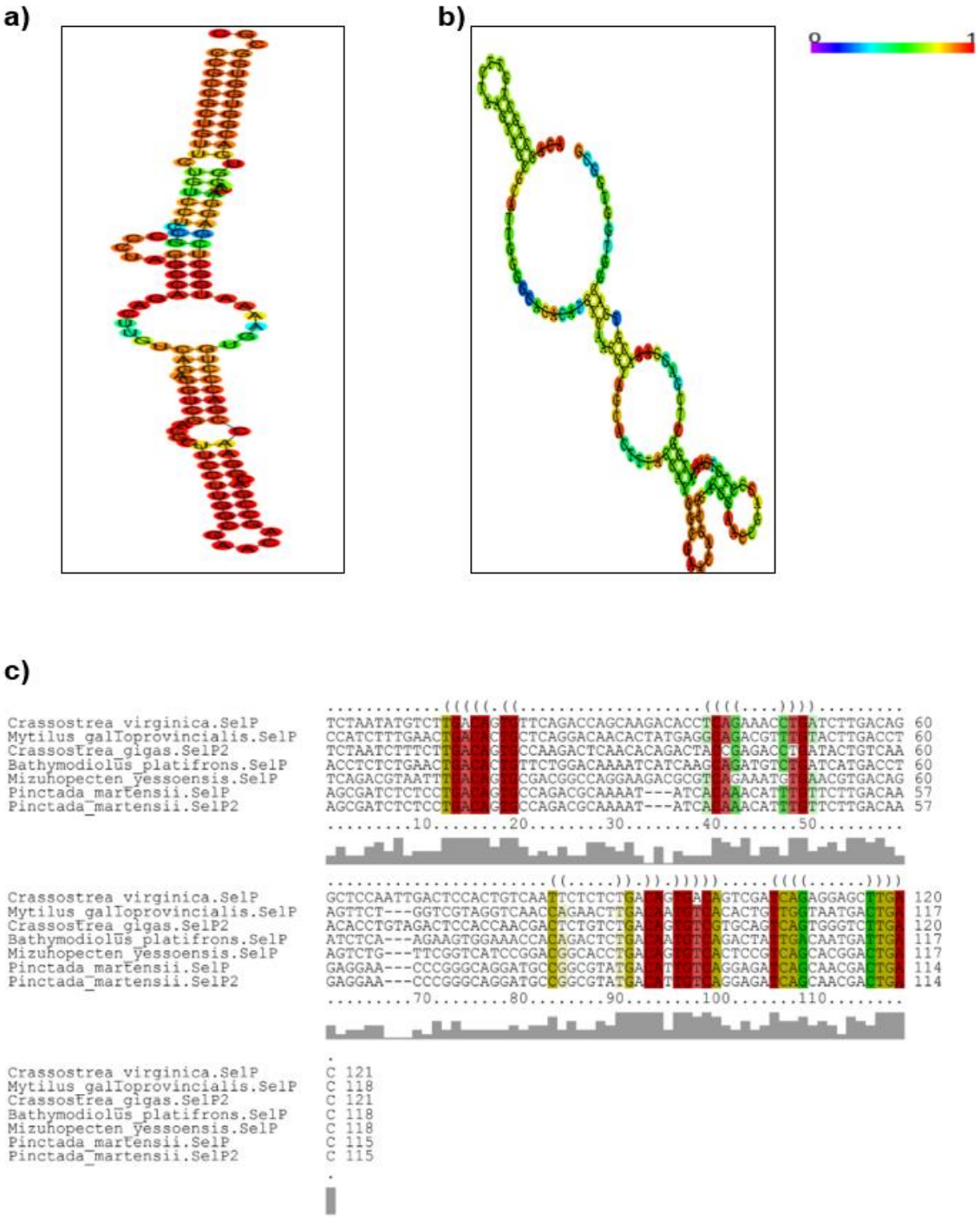


Supplementary Figure 3.5.1: HEK293T cell transfection and ^{75}Se labelling of invertebrate SelenoP constructs.
Autoradiography of HEK293T cells transfected with invertebrate plasmids (snail, spider, oyster exons only and oyster genomic DNA plasmids) from the media (a and b) and lysate (c) preparations. Blue arrow indicates endogenous HEK293T SELENOP secreted to the media and red arrow indicates zebrafish SELENOP full-length positive control.



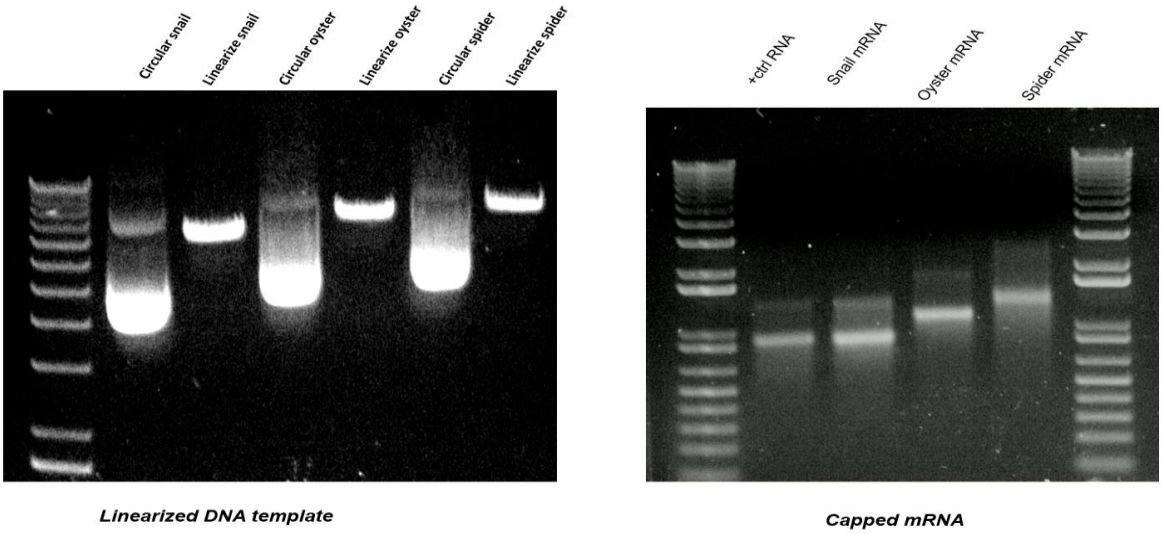
Supplementary Figure 3.5.2: Reconstitution of purified oyster SBP2 for zebrafish and rat SelenoP translation in RRL.

RRL reactions for (a) zebrafish (b) rat SelenoP mRNA reconstituted with increasing concentrations of full-length recombinant oyster SBP2 (GST-oyster SBP2-6XHis) or HRV-3C protease cut recombinant oyster SBP2 (Oyster SBP2-6XHis). A positive control of rat SelenoP mRNA reconstituted with rat CT-SBP2 loaded in a parallel lane (c) producing a full-length rat SelenoP radioactive product that resolves at 40 kDa. Black arrow points to single or double Sec incorporation events which is fainter than presented in Fig 3.3.9 when the cleaved oyster SBP2 was used.

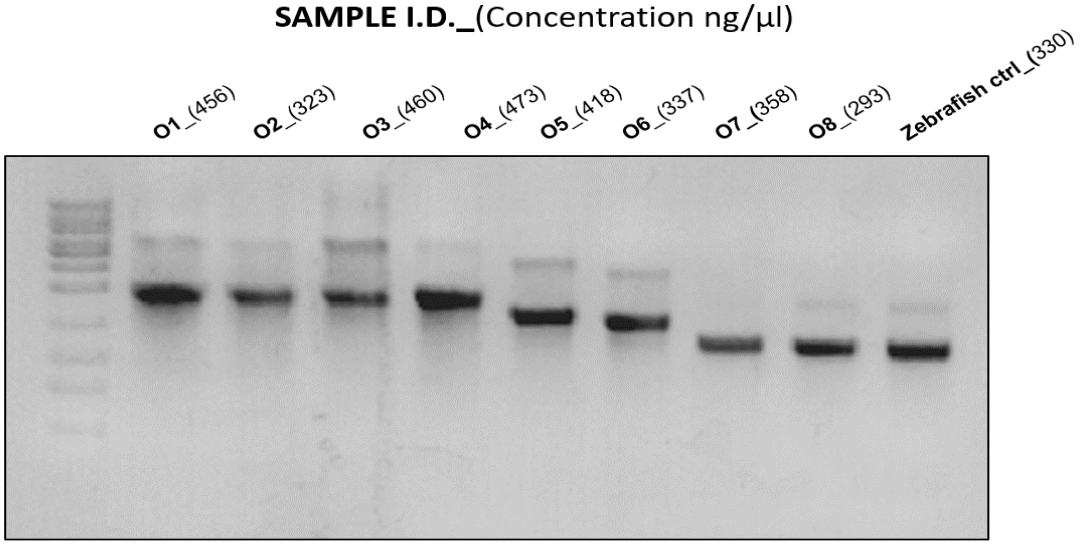


Supplementary Figure 3.5.3: Characterization of Initiation Stem Loop Structure (ISL) in bivalves.

(a)*C.gigas* SelenoP native ISL and (b) synonymously mutated ISL representation. The strength of complementarity is colour coded by the legend presented on the top right-hand corner with blue depicting the weakest and red the strongest sequence complementarity. (c) Multiple nucleotide sequence alignment of ISL in bivalves. Highlighted in colour is the nucleotide conservation across the six species listed.



Supplementary Figure 3.5.4: Quality control of invertebrate SelenoP DNA template and capped mRNA for RRL reactions
(a)Linearization of snail, spider and oyster SelenoP for subsequent T7 polymerase transcription. (b) mRNA integrity of the same constructs for RRL translation.



Supplementary Figure 3.5.5: RNA integrity gel for oyster mutant constructs
O1: native oyster construct, O2: Kozak consensus added., O3:5'UTR removed, O4: Initiation stem loop (ISL) disrupted. O5: C-terminal (UGA 11-46 deleted). O6: Oyster 3'UTR fused with zebrafish SelenoP O7: Oyster SECIS 1 fused with zebrafish SelenoP and O8: Oyster SECIS 2 fused with zebrafish SelenoP.

Chapter 4:

Probing *in-vivo* Selenoprotein P translation in *Crassostrea gigas* by Ribosome Profiling

Contributions:

Several contents of Chapter 4 are published (Baclaocos et al., 2019). To convey a complete description of the work undertaken the following contributions were included: Dr. Didac Santesmasses performed selenoprotein gene predictions, annotations and obtained RPKM/RPFKM values for differential gene expression. He also generated Table 4.1.1 and Figures 4.3.1, 4.3.3, 4.3.7 and 4.3.8.

- Total selenium determination by Inductively Coupled Mass-Spectrometry (ICP-MS) was performed by Dr. Katarzyna Bierla in University of Pau, France.
- All other experimental work and analysis in this Chapter was performed by the thesis author, unless otherwise stated.

4.1 Introduction

4.1.1: Background

Translation is tightly regulated in the cells through highly recognized mechanisms to vary the rates of initiation and elongation along the length of an mRNA (Sonenberg and Hinnebusch, 2009). For instance, ribosome access to the initiation codon can be affected by the length, sequence context and secondary structure of the 5'UTR. In some cases, the ribosome can bypass cap-dependent initiation and scanning through an internal ribosome entry site (IRES) (*Review* (Hinnebusch et al., 2016)). Slow decoding of some regions within an mRNA due to codon identity or coding sequence (CDS) context can lead to bottlenecks in translation evident by ribosome pausing which can occur for various regulatory purposes (Brule and Grayhack, 2017, Fletcher et al., 2000, Gardin et al., 2014, Wang et al., 2017). In addition, signals within an mRNA can mediate recoding events such as +1 or -1 frameshifting, translational bypassing, stop-codon readthrough and selenocysteine redefinition (Baranov et al., 2002a, Baranov et al., 2015).

While methods to measure global transcription regulation, like RNA-sequencing and micro-arrays have been well-developed, insights into the translational control of protein expression at a similar scale was lacking until the 2009 development of ribosome profiling (Ingolia et al., 2009). Ribosome profiling or ribo-seq is a relatively novel deep-sequencing method suitable to address questions regarding the global efficiency of protein expression *in vivo* and provide insights into the regulatory aspects of translation of an individual mRNA. Central to the ribosome profiling approach is the observation that a translating ribosome strongly protects a ~30 nucleotide mRNA fragment from ribonuclease activity (Steitz, 1969, Wolin and Walter, 1988).

Preparation of ribosome profiling libraries requires the collection of a biological sample and translation-inhibition to capture the ribosomes during the act of translation by flash-freezing or the use of drugs to inhibit elongation like cycloheximide. This is followed by ribonuclease digestion to release ribosome protected footprints (RPFs) and deep sequencing of cDNA libraries prepared from RPFs (Ingolia et al., 2012). Parallel library preparation to obtain a measure of mRNA abundance by RNA-sequencing is also performed to determine translation efficiency

(Ingolia et al., 2009). Bioinformatic pre-processing is then applied to the raw sequencing reads to exclude certain biases introduced during library preparation such as over-represented adapter sequences or bioinformatic removal of contaminating ribosomal RNA (rRNA). The processed sequencing reads are subsequently mapped back to a reference genome or transcriptome and a quality control step is performed to ascertain that the RPF fragments are derived from translating ribosomes (Fig 4.1.1). Characteristic distribution of RPFs include a high coverage in the coding sequence relative to 5'UTR and 3'UTRs, mRNA fragment size of ~30 nt and triplet-phasing corresponding to the codon step-size of the ribosome during translation.

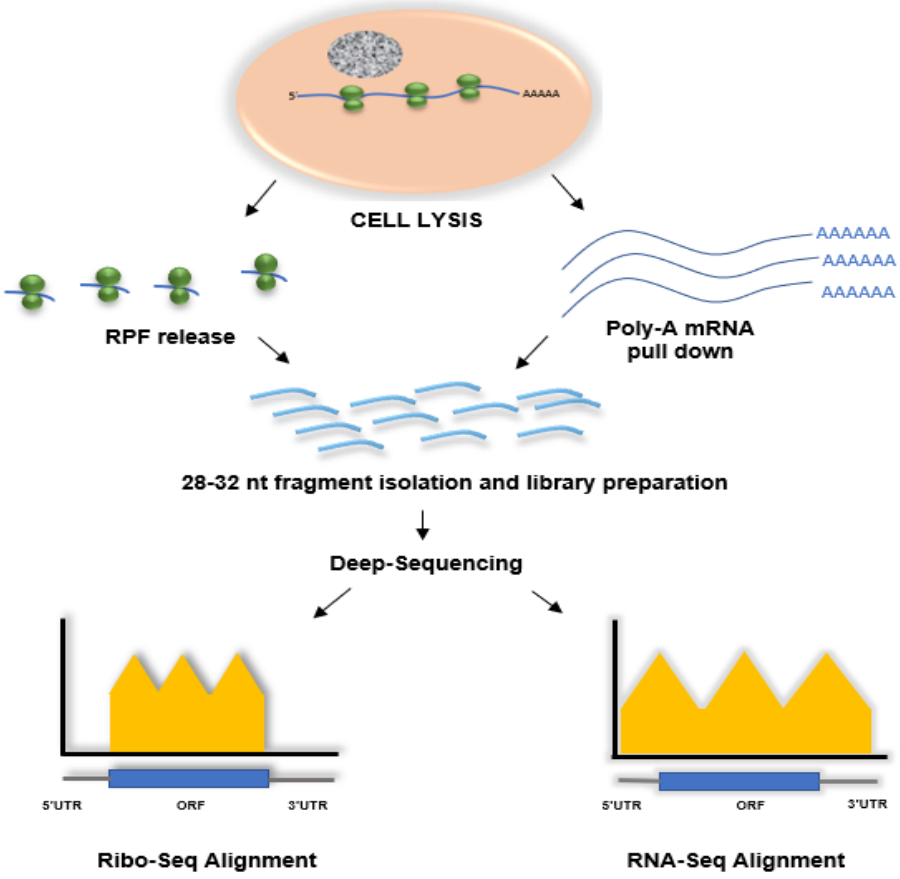


Figure 4.1.1: Schematics of ribosome profiling library preparation.
Biological samples are subjected to lysis followed by release of RPF fragments by nuclease digestion. Libraries are prepared from RPF fragments and from fragmented polyA mRNA isolated from total RNA. Ribosome density and mRNA abundance can be determined by mapping to a reference transcript.

The number of RPF counts reflects changes in ribosome density which can be applied for an individual mRNA shedding light into its mechanism of translation and

providing an approximate rate of protein synthesis. In addition, the positions of RPF distribution can also give insights into the mechanism of translational control such as alternative translation start sites, regulatory pauses marked by ribosome accumulation (Ingolia et al., 2011, Shalgi et al., 2013) and translated 5' open reading frame (Ingolia et al., 2014). Additional information such as initiation site, the frame being translated or whether the ribosome continues translation past a stop codon can be obtained from ribo-seq data leading to the discoveries of many novel or alternative protein isoforms (Brar et al., 2012, Stern-Ginossar et al., 2012, Ingolia et al., 2011, Dunn et al., 2013, Lee et al., 2012, Andreev et al., 2015).

Applications of ribosome profiling to study selenoprotein translation

Ribosome profiling is also a method suitable to study UGA-recoding, the efficiency of Sec incorporation and translational control during selenoprotein synthesis. Unique to selenoprotein translation is the requirement to redefine the UGA termination codon to Sec in a dynamic process involving interactions of cis-acting factors within selenoprotein mRNA (SECIS elements) and other factors of Sec-incorporation machinery (SBP2, eIF4a3, L30, eEFSec, $tRNA^{SerSec}$) (discussed in Chapter 1).

Studies investigating the efficiency of Sec-incorporation have used reporter constructs or cDNAs transfected into cultured cells. These experiments have consistently concluded that Sec-incorporation is inefficient and is highly dependent on Se regulation (Turanov et al., 2013, Banerjee et al., 2012). In selenoprotein P (SelenoP) mRNA where multiple UGA redefinition is required, Sec decoding was demonstrated to be inefficient at the first UGA in the 5' proximal region (Stoytcheva et al., 2006) and becomes highly processive upon Sec-incorporation in the distal UGAs (Fixsen and Howard, 2010). While studies in cell-culture or in cell-free translation systems have provided informative models of selenoprotein translation, measuring Sec-incorporation efficiency has not been previously possible due to the difficulty to measure an unstable termination product relative to full-length protein containing Sec (Lin et al., 2015).

To obtain a surrogate measure of Sec-redefinition efficiency, ribosome profiling can be used to monitor changes in ribosome density before and after Sec-UGA. RPFs from the annotated start codon (AUG) up to the Sec-UGA are designated

as 5' RPFs and are proportional to the ribosomes that have initiated translation on a selenoprotein mRNA. RPFs from Sec-UGA up to the subsequent stop-codon or 3' RPFs reflect the number of ribosomes that incorporated Sec and are translating downstream codons (Howard et al., 2013). Ribosome profiling has been used to confirm Se-dependent increase in translation of selenoproteins. An increase in RPFs after the Sec-UGA codon (3'RPFs) have been observed in mice fed with Se-supplemented diets compared to their Se-deficient counterparts (Howard et al., 2013). A kinetic feature of Sec-incorporation is the slow decoding of Sec-UGA evident by ribosomal pausing (Stoytcheva et al., 2006, Fletcher et al., 2000). RPF accumulation observed upstream of Sec-UGA suggests that a delay in translation occurs prior to the ribosome encountering the Sec-UGA codon. This is postulated to be a regulatory control mechanism for the recruitment of factors needed to incorporate Sec. More recently, ribosome profiling has been used to monitor selenoprotein synthesis *in-vivo* in gene knock-out or gene edited mice (Fradejas-Villar et al., 2017, Mariotti et al., 2017) to gain insights into the regulation and mechanisms involved during selenoprotein translation.

4.1.2: Chapter Aims

The high number of Sec residues in the C-terminal domain of *C.gigas* suggests an extremely efficient mechanism of Sec- insertion upon translation. We have previously shown that a heterologous vertebrate RRL was unable to support full-length processive Sec incorporation in oyster SelenoP distal UGAs and can only generate oyster SELENOP termination products (Chapter 3). Synthesis of such termination products from the oyster SelenoP and incompatible oyster SBP2 interaction, indicates insensitivity of the vertebrate translation components to signals in oyster SelenoP mRNA and/or unknown oyster trans-acting components lacking in RRL, necessary for processive Sec incorporation. Since SelenoP translation *in-vitro* was shown to be inefficient, we aimed to characterize its *in-vivo* translation by ribosome profiling of oysters cultivated in natural and Se-supplemented conditions. We also aim to confirm SelenoP metabolic Sec-incorporation *in-vivo* by ^{75}Se labelling. The rich data set generated from ribosome profiling allowed us to explore, in addition, the general effects of selenium on selenoprotein mRNA and protein abundance as well as for Sec-UGA redefinition. The implication of selenium supplementation on oyster

selenoprotein expression and regulation will be discussed as well as the mechanistic insights gained on SelenoP translation from our analysis.

4.2 Materials and Methods

4.2.1 Tetraselmis sp. algal culture

Artificial sea water (ASW) made with Instant Ocean Aquarium sea salts (Instant Ocean) adjusted to a salinity of 35 ppt was used to culture *Tetraselmis sp.* obtained from an oyster cultivation farm were grown for oyster feed. Algal cells were diluted in 80 ml of algae media (80% ASW and 20% F2 Guillards media (Sigma G0154)), either supplemented or unsupplemented with 5 mg/L of sodium selenite (11.28 µM) (Sigma- 214485) diluted in artificial sea water. Algae were grown for 3 days with constant aeration in a 12-hour dark and 12-hour light cycle to allow for growth and selenium incorporation. On the day of feeding, cells were equalized by cell numbers, counted under a light microscope, and fed to each corresponding oyster tanks weekly.

4.2.2 Aquaculture and selenium supplementation of C.gigas

Two- year old diploid Pacific oysters obtained from an oyster farm in Galway, Ireland were acclimated into sea-water for one week. After a week, they were distributed equally into tanks (10 per tank) containing 10L volume of artificial seawater with salinity adjusted to 25 ppt and kept at constant temperature of 16°C. The tanks were constantly aerated and were kept at a 12-hour light/dark cycles. Nitrifying bacteria (obtained from a local pet-store) were added during the first set up to help establish a filtration system. Each oyster tank was fed weekly for 6 weeks by pouring the corresponding algal feed grown in the presence or absence of selenium. To avoid oyster mortality caused by a build-up of waste, ASW was changed every two weeks and nitrifying bacteria was topped up weekly.

4.2.3 Gonad histology and sex determination

Twenty-four hours after the last feeding day, the oysters were sacrificed, and tissue cross-sections were prepared for histology to determine sex and gonad development stage. A cross section of the gonad tissue (5mm thickness) was placed in a histology cassette (Sigma) and stored in 100% ethanol for 24h. The dehydrated tissue was fixed with 90% fixative (Davidson Solution-Sigma) and 10% acetic acid for 24-48 hrs. A second fixing step followed by clearing and a paraffin embedding steps were

performed. The paraffin penetrated tissue was embedded using hot paraffin wax and allowed to solidify. The paraffin embedded tissue was prepared for staining by slicing 7µm sections using a microtome Leica RM2235 (Leica). Cut paraffin-embedded sections were placed in a warm 30°C water bath and placed into glass slides. These were allowed to dry for 24h. The dried slides were stained in cycles of ethanol, acetic-acid, haemotoxylin, eosin and DPX mounting medium. The slides were analysed under a light microscope at 40X magnification and images was captured with Hamamatsu ORCA-ER Digital Camera C4742-80 and processed using Andor IQ acquisition software (Andor Technology Ltd. Belfast, Northern Ireland). Only male oysters with developed gonads were subjected to total selenium determination and ribosome profiling library preparation.

4.2.4 Inductively Coupled Mass Spectrometry (ICP-MS)

Whole-body male oyster tissues were prepared for ICP-MS by liquid nitrogen grinding. Independently, the algae-feed, tank water, and oyster cytoplasmic extract preparations were also measured by the following procedure. Chromatographic separations were carried out using a Model 1100 or 1200 HPLC pump (Agilent, Wilmington, DE, USA) as the delivery system. The exit of the column was directly connected to the Meinhard nebulizer (Glass Expansion, Romainmotier, Switzerland) of the ICP MS equipped with a collision cell (Agilent 7700, Tokyo, Japan) by means of PEEK tubing. Injections (both in size-exclusion) were performed using a Rheodyne valve with a 100-µL sample loop. DigiPrep (SCP Science, Courtaboeuf, France) was used to heat the sample during acid digestion.

The samples were weighed and left overnight in 0.2 to 0.5 mL of HNO3 (depending on quantity of the sample available). One mL of H_2O_2 was added and the sample was digested in a DigiPrep (the digestion program: 0-30 min from room temperature up to 65 °C, 30-240 min - kept at 65 °C). The digests were diluted to reach the HNO_3 concentration of 4 % and analyzed by ICP MS using the optimized daily conditions (nebulizer gas flow, RF power, lens voltage and collision gas flow) (Ogra et al., 2004). We carried out external calibration at 6 levels, adding selenium to a blank sample (mixture of nitric acid, hydrogen peroxide and water) to reach Se concentrations of 0.25, 0.5, 1, 2.5, 5 and 10 ppm.

4.2.5 Ribosome Protected Fragments Preparation

A. Sucrose gradient preparation

For ribosome profiling library preparation, continuous sucrose gradients (10% or 60% sucrose, 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$, 1mM DTT and 100 µg/ml cycloheximide) were made by transferring 60% sucrose solution (5 ml) into ultra-centrifuge tubes (Beckman Coulture) followed by careful layering of an equal amount of 10% sucrose solution. The tubes containing sucrose solutions were carefully tilted to a 90° right angle and allowed to form continuous gradients for 4 h at room temperature. The sucrose gradients were carefully placed upright and stored frozen at -20°C until needed.

B. Tissue lysate preparation

Tissue lysates were prepared from two individual whole oyster soft tissues that had similar determined levels of selenium from each supplementation group. These were pooled and ground in liquid nitrogen using pestle and mortar and lysed in 2 ml polysome lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 15 mM $MgCl_2$, 1% sodium deoxycholate, 1% Triton X-100, 100µg/ml cycloheximide, 2 mM DTT and 2µl SuperaseIN (Thermofischer). Lysates were centrifuged at 12 000 x g for 10 min in 4°C.

C. Ribosome protected fragment (RPF) isolation

One ml of cleared lysate was digested with 100U of RNAse 1 (Thermo Scientific) at 25°C for 45 minutes and in parallel non RNAse-1 treated lysates were also used to determine intact polysomes. Monosome fractions were isolated by loading RNAse-1 treated lysates on the prepared continuous sucrose gradient followed by centrifugation at 100 000 x g for 3h at 4°C. Gradients were chased with 80% caesium chloride prepared with bromophenol blue dye in a gradient fractionator (Brandel Density Gradient Fractionation System). Fractions were collected in a 96-well plate at 12s intervals and monosome fractions containing RPFs were determined by absorbance reading at OD_{256}. Collected monosome fractions were pooled and RNA was extracted using Trizol LS (Ambion) as per manufacturer's protocol. RPFs were isolated by 15% PAGE-Urea gel purification as outlined in method section 4.2.8. The purified RPF RNA was denatured and prepared for end repair and 3' linker-ligation.

4.2.6 Poly-adenylated (poly-A) mRNA preparation and alkaline hydrolysis

Total RNA (100 µg) was extracted from total tissue lysate preparation by Trizol-LS (Ambion) as per manufacturer's protocol. Poly-A mRNA was isolated using polyA purist magnetic kit (Thermo Fischer) as per manufacturer's guideline. The resulting poly A mRNA was incubated with 1 volume of alkaline fragmentation buffer (2mM EDTA, 10 mM Na_2CO_3, 90 mM $NaHCO_3$) for 55 min at 95°C to randomly cleave mRNA. The fragmented RNA sample was precipitated over-night. The recovered RNA was PAGE-purified from a 15% TBE-Urea gel (see Method 4.2.8) while the 28-32 nt RNA ladder was used to guide band excision of the same size mRNA fragmented band. RNA was extracted from the gel and precipitated for end-repair and 3' linker ligation

4.2.7 cDNA library preparation for deep sequencing

The subsequent steps of library generation used the isolated RNA prepared from the appropriate preceding step as a template and described in order from A-D. RNA/DNA purification from PAGE gels were performed using method sections 4.2.8 and 4.2.9.

A. RNA End-Repair and 3' Linker Ligation

Denatured RNA (100 ng in 10µl) from RPF isolation and mRNA fragmentation were end-repaired by addition of 1µl of T4 Polynucleotide Kinase (PNK) (NEB), 5µl of 10X PNK buffer (NEB), 1µl SUPERase IN (Invitrogen) to each reaction. The reaction mix was incubated for 1h at 37°C and PNK was heat-inactivated at 70°C for 10 min. RNA was precipitated at -80°C for 1h by addition of 50µl water, 10µl 3M NaOAc, 2 µl glycoblue and 150µl isopropanol. RNA was recovered by centrifugation at 14,000 x g at 4°C, followed by a 1 ml 75% Ethanol (Sigma) wash. The RNA pellet was air-dried and subsequently resuspended in 10µl of milli-Q water. To 10 µl of precipitated RNA resuspended in milli-Q water, 1µl of pre-adenylated linker (IDT) was added. The reaction was denatured and a linker ligation reaction (2µl 10X T4 RNA ligase 2-truncated (Rnl2) ligation buffer (NEB), 1µl T4 Rnl2 ligase (NEB), 1µl SUPERaseIn and 6 µl 50% Poly-ethylene Glycol 8000 (PEG8000) (NEB)) was prepared. The samples were incubated overnight at 4°C and the RNA precipitated the next day. Following precipitation, the ligated RNA was PAGE purified from a 15% Urea-TBE gel and extracted for reverse transcription.

Pre-Adenylated Linker Sequence	1/5rApp/CTGTAGGCACCATCAAT/3ddC/

B. Reverse transcription

The ligated RNA (10µl) was reverse transcribed using Superscript III kit (Invitrogen). The reverse transcription (RT) primer (1.25µm) containing sequences complementary to the linker was added into the ligated RNA and denatured for 2 min at 80°C. The kit components were added as per manufacturer's guideline. RT reactions were carried out for 30 min at 48°C followed by RNA hydrolysis with 2.2µl of 1M NaOH for 20 min at 95°C. First strand cDNA was precipitated and PAGE-purified by resolving in to a 7.5% TBE-Urea gel along with the preparation of the RT primer on its own as a negative control. Reverse transcribed First-strand complementary DNA (cDNA) was extracted and prepared for circularization.

RT-Primer Sequence 1341	**5′-(Phos)-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGC-(SpC18)-CACTCA-(SpC18)-TTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCTACAG-3′**

C. Circularization

First strand cDNA was circularized by addition of the following kit components: 2 µl 10X Circligase II Buffer, 1 µl 50 mM $MnCl_2$, 1 µl circligase enzyme (Epicentre) to 16µl cDNA. The reaction was incubated at 60°C for 2 hr and the enzyme was heat-inactivated at 80°C for 10 min. The circularized DNA was stored in -20°C until required.

D. Polymerase chain reaction (PCR) for library preparation

A PCR reaction (100µl) was prepared for each sample for a final library amplification. Components of PCR reaction include: 20µl 5X High Fidelity Phusion buffer (NEB), 2µl circularized PCR template, 1.6µl 10mM dNTPs (Solis Biodyne), 5µl lllumina forward primer (1350), 5µl lllumina reverse primer containing the sequence index for multiplex sequencing (1342, 1347, 1344, 1345), 1µl Phusion polymerase (NEB) and 64.5µl milli-Q water.

The template was amplified for 8-16 cycles using the following cycle settings:

Cycle	Temperature (°C)	Time (s)
Cycle 1: Initial Denaturation	98	30
Cycles 2-16: Denaturation	98	10
Primer Annealing	65	10
Elongation	72	5
Final elongation	72	10
∆Ct 0.1°C/sec (for steps 1-4)		

Table 4.2.1 Illumina primer forward and reverse sequences for final cDNA library amplification

Primer I.D.	Sequence (5'-3')	Index (barcode)	Sample Description
Forward **1350**	AATGATACGGCGACCACCGAGATCTACAC		
Reverse			
1464	CAAGCAGAAGACGGCATACGAGATTGGTCA GTGACTGGAGTTCAGACGTGTGCTCTTCCG	TGACCA	Non-Supplemented RNA-seq
1463	CAAGCAGAAGACGGCATACGAGATGCCTAA GTGACTGGAGTTCAGACGTGTGCTCTTCCG	TTAGGC	Non-supplemented Ribo-seq
1345	CAAGCAGAAGACGGCATACGAGATACGTCG GTGACTGGAGTTCAGACGTGTGCTCTTCCG	CGACGT	Se Supplemented RNA-seq
1344	CAAGCAGAAGACGGCATACGAGATATGCTG GTGACTGGAGTTCAGACGTGTGCTCTTCCG	CAGCAT	Se Supplemented Ribo-Seq

4.2.8 Polyacrylamide Gel Electrophoresis (PAGE) gel purification

To purify appropriate RNA templates, a 15% (for RNA samples) or 7.5 % (for cDNA samples) denaturing Urea-TBE gel was poured and allowed to solidify for 30 minutes. This was pre-ran at 15 mA for 30 minutes in 1X TBE buffer. RNA loading dye was added to extracted RNA followed by denaturation for 90s at 80°C. The denatured

RNA/cDNA was resolved for 60-90 min with a parallel appropriate ladder to guide the excision of the nucleic acid band of interest. The gel was stained with SYBR-Gold (Thermo-Fischer) diluted 1:20 with 1X TBE buffer for 1 min. The band of interest was visualized under a UV-light, excised and extracted. A parallel ladder was prepared similarly as a positive control. To isolate the final cDNA library, a non-denaturing TBE polyacrylamide gel (8%) was prepared. DNA-loading dye was added directly to circularized cDNA. Samples were dissolved for 60 min. The cDNA was extracted from the gel as described.

Table 4.2.1: Recipe for polyacrylamide TBE-Urea gels

Gel Components	Polyacrylamide gels (Volume)		
	15%	**7.5%**	**8%**
40% acrylamide/bis-acrylamide (19:1)	5.63 ml	2.82 ml	3 ml
Urea	7.2 g	7.2 g	-
10X TBE	1.5 ml	1.5 ml	1.5 ml
Water	1.9 ml	4.72 ml	10.5 ml
37°C to dissolve then filter before adding			
10% APS (ammonium persulfate)	37.5 µl	37.5 µl	37.5 µl
TEMED	7.5 µl	7.5 µl	7.5 µl

4.2.9 RNA and cDNA extraction and precipitation

For each corresponding step of cDNA library preparation, RNA and cDNA were extracted from the excised band by the addition of nucleic acid extraction buffer (300 mM sodium acetate (NAOAc) pH 5.3, 1mM EDTA and 0.1% SDS), flash-freezing and overnight shaking at room temperature.

RNA or DNA were recovered from the extraction buffer by addition of 2 µl GlycoBlue (Invitrogen) and 1 volume of isopropanol followed by freezing at -80°C for at least 1h. The samples were centrifuged at 14,000 x g for 30 min. The blue pellet was washed with 80% ethanol and centrifuged at 14,000 x g for 20 min. It was allowed to dry and resuspended an appropriate volume of water. RNA or DNA concentrations were determined using a Qubit Assay (Invitrogen) as per manufacturer's guideline.

4.2.10 Library sequencing

The prepared cDNA libraries were sent for sequencing to a commercial service offered by BGI Genomics Co., Ltd. Hongkong (www.bgi.com). The samples were pooled in one SE50 multiplex lane where 3GB of data was allocated for each ribosome profiling samples and 0.6 GB of data was allocated for RNA-sequencing samples.

4.2.11 Application of bioinformatic filters to raw sequencing reads and subsequent alignment

Raw sequencing reads were pre-processed before alignment using available online tools in ribogalaxy (ribogalaxy.ucc.ie) (Mullan et al., 2016). The raw reads obtained per sample were subjected to FASTA QC followed by 3'adapter sequence removal that is expected to be over-represented in all sequence reads. After adapter trimming, ribosomal RNA (rRNA) sequences were bioinformatically removed using 5S (ENSRNA022717831), 5.8S (ENSRNA02271792), 18S (ENSRNA022718259) and 28S (AB102757.1) *C.gigas* rRNA sequences obtained from Ensembl. The remaining sequences were mapped to the full oyster transcriptome to generate a metagene analysis plot for quality control followed by mapping to annotated selenotranscriptome and selenoprotein machinery.

4.2.12 Annotation of C.gigas selenoproteins and selenoprotein machinery

Selenoprotein and machinery transcript sequences were obtained from a comprehensive transcriptome assembly (Riviere et al., 2015) and the published available genome from Ensembl. Selenoprotein predictions were made using Selenoprofiles (Mariotti and Guigo, 2010).

4.2.13 Calculations to obtain differential expression between samples

Reads per kilobase pair of transcript length per million mapped reads (RPKM) determination

To calculate mRNA abundance and ribosome foot-print coverage of selenoprotein genes and machinery, read counts were obtained along the coding sequence of each gene. The reads were normalized by reads per kilobase (kb) length of transcript per million mapped reads (RPKM or RPFKM). RPFs and mRNA abundance of the non-supplemented and supplemented samples were plotted as log (RPKM)+1.

A. Translation efficiency

Translation efficiency was calculated by normalizing RPFKM from ribosome profiling with RPKM from RNA-sequencing (RPF/RNA).

B. UGA redefinition efficiency calculation

To calculate UGA redefinition efficiency, RPKMs from the 5' length of the gene including AUG to Sec-UGA and 3' length of the gene including Sec-UGA to the real stop codon were obtained separately. Reads with the A-site mapping to the first and last 15 nts of the annotated coding sequence were excluded to avoid bias at the initiation and termination codon. Analysis of RPFs upstream of Sec-UGA codons excluded UGA and the five preceding codons (Howard et al., 2013, Fradejas-Villar et al., 2017).

4.2.14 Ribosome-coverage map generation

Data-set alignments, along with a reference transcriptome were uploaded to trips-viz (trips.ucc.ie) (Kiniry et al., 2019) to visualize ribosome abundance and RNA coverage across individual selenoprotein mRNA. Plots were generated for selenium supplemented and non-supplemented sample for SelenoP. Mapping of ambiguous reads were allowed due to the presence of a second isoform. Minimum and maximum fragment length were set at 25 to 35 nt respectively. Images were generated and downloaded from the trips-viz website.

4.2.15 Oyster larvae in-vivo ^{75}Se labelling

Oyster larvae (approximately 100,000) at 7 and 14-day old stages were obtained as a gift from the Pacific Sea Foods, Quilcene, Washington . After 24h acclimation in 10L of artificial sea water and initial feeding with Instant Shellfish Diet1800, 100-200 oyster larvae (100 µm in size) were transferred into a 12 well plate with artificial sea water spiked with 1 µl of 100 µM ^{75}Se diluted in 1% DMSO for metabolic selenium incorporation. After a 24h timepoint, oyster larvae were washed, harvested and mechanically lysed in RIP-A lysis buffer (50 mM Tris-HCl pH 8.0, Triton X-100, 150 mM NaCl, 0.5% sodium deoxycholate and protease inhibitors). Protein extracts were electrophoresed in 10% SDS-page gel, fixed, dried and exposed to a phosporimager screen. Parallel Western blot probing was performed.

4.2.16 Custom Antibodies Generation and Protein Analysis

Custom antibodies for oyster SELENOP have the immunogenic peptide sequences TADGTDPVKARVN (N-terminal) and YCRTGTYDDRAH (C-terminal). These were generated similarly as in (2.2.2). For SELENOP immunoblot analysis of oyster lysates, samples were equalized to 50 µg per lane, resolved in a 12% SDS-PAGE gel and transferred to nitrocellulose membrane (Protran). The membrane was blocked with 5 % milk in phosphate buffer saline-Tween20 (0.5% PBS-T) over-night in 4°C and probed with 1 :2500 dilutions of rabbit primary antibodies (anti-NT-SELENOP) and (anti-CT-SELENOP) in 5% milk-PBS-T for 1 hr at RT. The membrane was washed 3 times with PBS-T. Immunoreactive bands were detected with appropriate fluorescently labelled secondary antibodies and scanned using LI-COR Odyssey® Infrared Imaging Scanner (LI-COR Biosciences). For comparison of SELENOP abundance, full-length product band intensity was quantified using Image Lite Studio. Values were normalized against Ponceau S loading control and plotted as an average of two technical replicates.

4.3: Results

Oysters play a vital role in estuarine and coastal marine habitats and have developed many physical adaptations for survival in a highly dynamic and stressful environment. Pacific oysters are economically important bivalves, cultivated for human consumption. As well as the remarkably high numbers of Sec-UGA in its SelenoP mRNA which is suggestive of a unique genetic decoding mechanism, there are secondary implications of selenium uptake and utilization indicative of elevated Se levels identified in its natural environment (Zhang and Gladyshev, 2008). A more extensive analysis on the Pacific oyster selenoproteome was performed, using ribosome profiling, providing insights into the effects of selenium levels on the transcription and translation of its selenoprotein genes.

4.3.1 The Pacific oyster selenoproteome

Thirty-one selenoprotein genes were identified in the genome sequence of *C.gigas* along with a complete set of factors required for selenoprotein synthesis (tRNA-Sec, pstk, SecS, SBP2, eEFSec and SPS2). These selenoproteins belong to 22 distinct protein families with very similar composition to the protein families previously reported in an evolutionary analysis of the metazoan selenoproteome (Jiang et al., 2012). The *C.gigas* genome encodes at least one selenoprotein gene for each of the 21 families described in that study. Three of those are not found in vertebrates namely AhpC, MsrA and DsbA. All selenoprotein genes and Sec-translation factor genes were also identified in a comprehensive transcriptome assembly (Riviere et al., 2015), providing evidence for their transcription (Table 4.3.1).

	Protein Family	Genome	Transcriptome
Selenoproteins	AhpC	JH816908.1	CHOYP_SORBIDRAFT_01G038790.1.1
	DI	JH816160.1	CHOYP_DIO1.1.1 CHOYP_IOD1.1.2
	DI	JH817465.1	CHOYP_IOD1.2.2
	DsbA	JH818180.1	CHOYP_LOC100375622.1.1
	GPx	JH815978.1	CHOYP_GPX1.1.4
	GPx	JH816450.1	CHOYP_GPX5.1.1
	GPx	JH816545.1	CHOYP_GPX1.3.4 CHOYP_GPX4-A.1.1 CHOYP_GPX1.4.4
	GPx	JH818884.1	CHOYP_GPX3.1.2
	MSRB1	JH816382.1	CHOYP_MSRB1.1.2 CHOYP_MSRB1.2.2
	MsrA	JH816751.1	CHOYP_PHUM_PHUM280830.1.1
	Rsam	JH817292.1	CHOYP_BRAFLDRAFT_118756.1.2
	SELENOF	JH816163.1	CHOYP_SEP15.1.1
	SELENOH	JH818600.1	CHOYP_SELH.1.1
	SELENOJ	JH816288.1	CHOYP_LOC101155972.1.1 CHOYP_SELJ.1.1
	SELENOK	JH816435.1	CHOYP_LOC580102.2.3 CHOYP_BRAFLDRAFT_118197.1.1 CHOYP_LOC580102.1.3 CHOYP_LOC580102.3.3
	SELENOL	JH817326.1	CHOYP_SELL.1.1
	SELENOM	JH816887.1	CHOYP_LOC590298.3.3 CHOYP_SELM.1.3 CHOYP_SELM.2.3
	SELENON	JH817706.1	CHOYP_SELN.1.1
	SELENON	JH818298.1	CHOYP_LOC100709538.1.2
	SELENOO	JH817782.1	CHOYP_SELO.1.1
	SELENOP	JH816366.1	CHOYP_SEPP1.2.3
	SELENOP	JH822951.1	CHOYP_SEPP1.3.3 CHOYP_LOC100304446.1.2 CHOYP_LOC100871266.1.1
	SELENOS	JH818024.1	CHOYP_TSP_04780.1.1 CHOYP_LOC100376740.1.1
	SELENOT	JH817380.1	CHOYP_LOC100368444.1.2 CHOYP_LOC755635.1.1
	SELENOU	JH816091.1	CHOYP_F213A.1.1
	SELENOU	JH817465.1	CHOYP_RS6.7.12
	SELENOU	JH821057.1	CHOYP_BRAFLDRAFT_96268.1.1
	SELENOW	JH817593.1	CHOYP_SEPW2B.2.2 CHOYP_SEPW2B.1.2
	SPS2	JH817936.1	CHOYP_SELD.1.2 CHOYP_SELD.2.2 CHOYP_NEMVEDRAFT_V1G135670.1.1 CHOYP_HSP7D.3.3
	TR	JH816091.1	CHOYP_BRAFLDRAFT_122807.1.1
	TR	JH822994.1	CHOYP_TRXR2.1.1
Machinery	SBP2	JH816448.1	CHOYP_LOC584318.1.1 CHOYP_SECISBP2L.1.1 CHOYP_TBCD.2.2
	SecS	JH816140.1	CHOYP_SPCS.1.1
	eEFsec	JH818980.1	CHOYP_LOC100375943.1.1
	Pstk	JH823230.1	CHOYP_LOC100699733.1.1

Table 4.3.1*: The C.gigas* selenoproteome

The protein family, corresponding contig in the genome, corresponding transcripts in the transcriptome assembly for the 31 selenoprotein genes and the Sec machinery in *C.gigas* are indicated.

C. gigas also encodes a Sec-containing Radical S-adenosyl methionine (RSAM) protein (Radical SAM/Cys-rich domain; Interpro IPR026351). RSAM are enzymes that generate a radical species by reductive cleavage of S-adenosyl methionine. It catalyses diverse reactions such as methylation, isomerization, sulfur insertions, and oxidation (Shibata and Toraya, 2015). RSAM selenoproteins have never been observed in metazoans but have been previously reported in bacteria and in a single-cell eukaryote of the harmful bloom alga *Aureococcus anophagefferens* (UniProt F0XY08) (Gobler et al., 2013). The gene encoding RSAM in the oyster has a eukaryotic SECIS in the expected location. We also found evidence for its expression in the *C.gigas* transcriptome. A search in the GenBank Transcriptome Shotgun Assembly (TSA) database revealed numerous RSAM selenoproteins in other bivalves and metazoan lineages. Crustaceans, bivalves and most cnidarian species contain Sec in their RSAM protein while in most echinoderms, the cysteine homologue of RSAM is present. (Fig. 4.3.1).

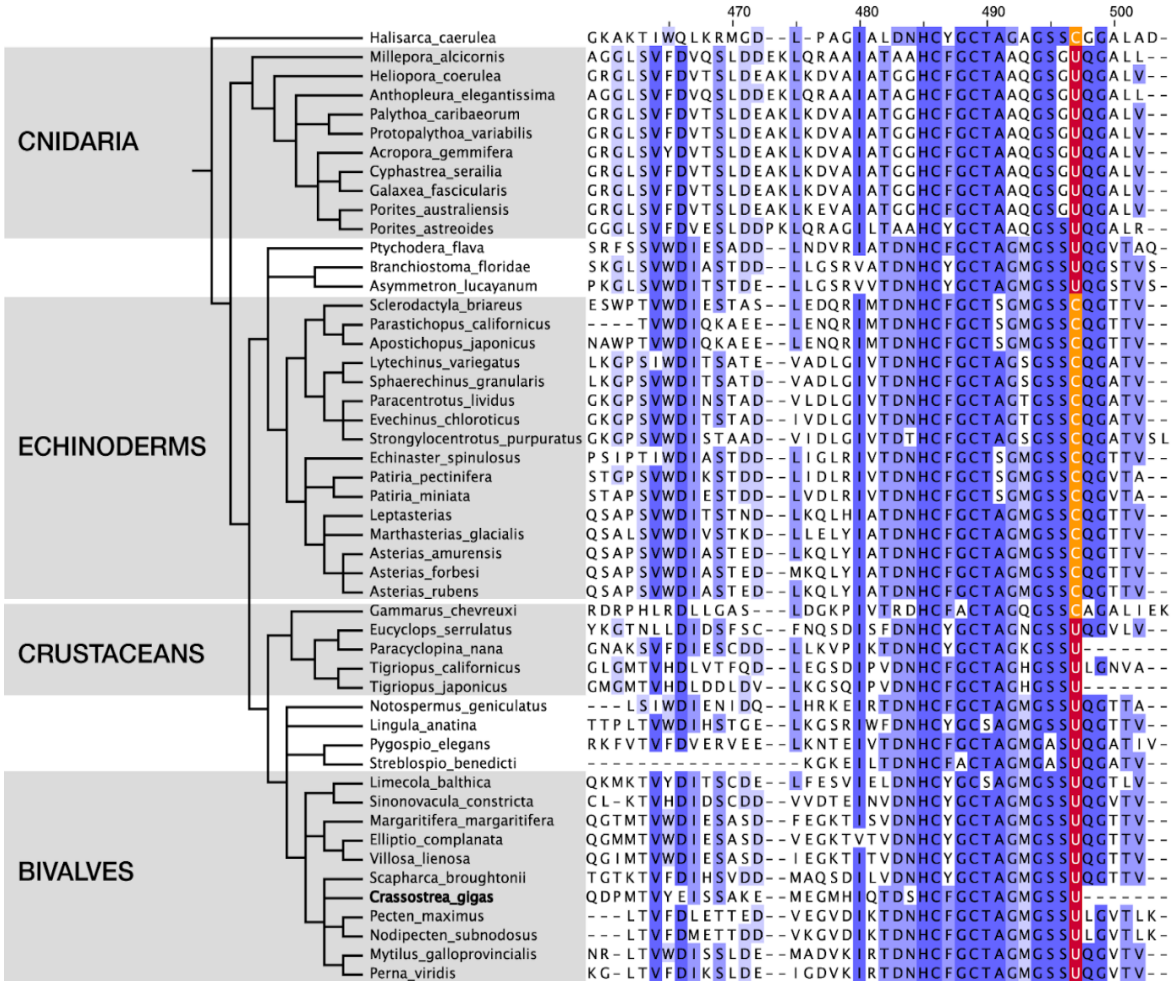


Figure 4.3.1: Radical S-adenosyl Methionine (RSAM) selenoprotein in *C.gigas* and other species. Multiple amino acid sequence alignment of Radical SAM (RSAM) proteins, showing only the C-terminus including the Sec position (column 497). The residue at the Sec position is highlighted in red for Sec and in orange for Cys. The sequences were obtained from NCBI GenBank (TSA). The phylogenetic tree on the left was obtained from NCBI Taxonomy.

4.3.2 Selenium uptake and accumulation in oyster tissues

Oysters feed on natural phytoplankton by filter-feeding. They are known for their great capacity to filter microscopic food source from water as well as their ability to bioaccumulate sediments, nutrients and even pollutants from their environment (Wang et al., 2018). To investigate the effects of selenium on the *C.gigas* selenoproteome oyster selenium supplementation trial was performed where their microalgae food source (*Tetraselmis sp*.) was grown in Se-rich medium by the addition of 11.82 µM (5mg/L) final concentration of inorganic sodium selenite. This is the concentration previously suggested in reported studies that will allow optimal Se- uptake without inducing toxicity to the algal cells (Umysova et al., 2009, Sun et al., 2014).

Mammalian brain and testes preferentially take up long forms of SELENOP from plasma (Kurokawa et al., 2014). Analysis of publicly available RNA-seq datasets revealed higher expression levels of SelenoP RNA in fully developed male oysters compared to both their female counterparts, and oysters at earlier stages of development (Riviere et al., 2015, Zhang et al., 2012). To study high level SelenoP translation, adult male oysters with developed gonads were utilized, identified through nuclei-staining of male gametes and mature follicle sacs from histological preparations (Fig. 4.3.2 a(ii)). Two groups of 10 oysters were distributed to separate tanks and fed weekly for 6 weeks with equal amounts of micro-algae pre-grown with or without Se-supplementation.

Total selenium levels in tissues of selenium supplemented oysters increased 20-fold on average compared to non-supplemented controls (Fig. 4.3.2 b). Without mortality, individual male oysters accumulated 50-fold higher selenium levels than in the control group (Table 4.3.2). As also analyzed by ICP-MS (Methods), other material from the supplemented tank, oyster non cytoplasmic debris, sea water, and the algae that served as food source, also showed elevated selenium levels (Table 4.3.3).

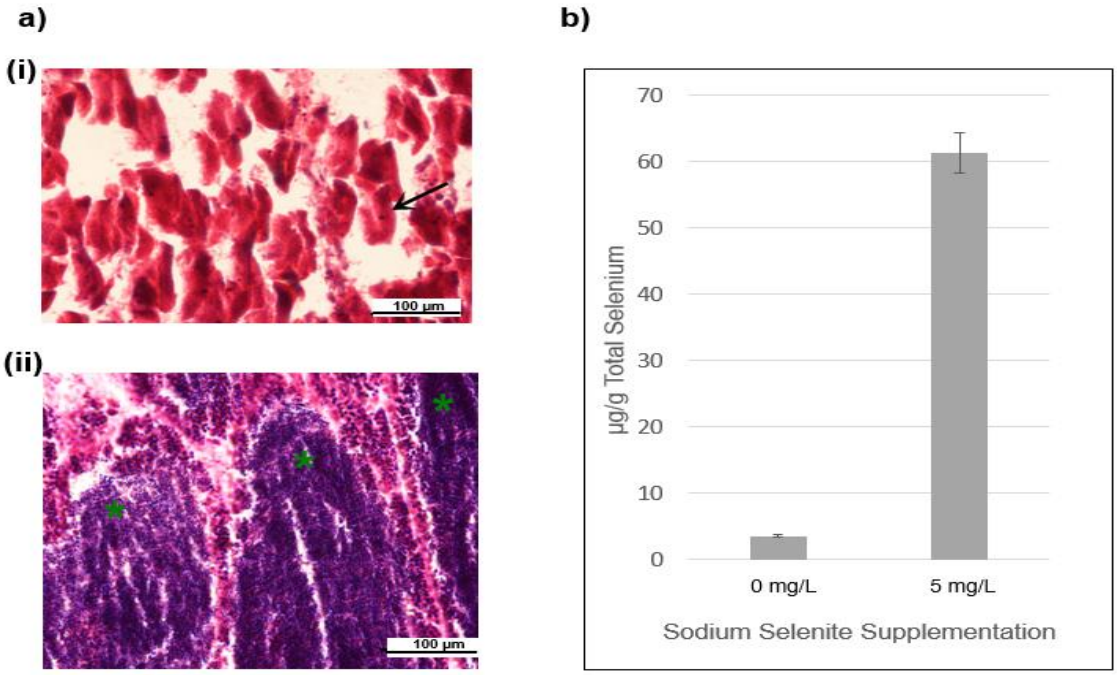


Figure 4.3.2: Characterization of oyster tissues after selenium supplementation.
(a)Histology of oyster gonads stained with haematoxylin and eosin stains. (i) Female gonad is stained with varying shades of red/pink. The arrow points to the blue-black stain of the nuclei of primary oocytes (ii) male oyster gonad shows purple staining of sperm cells nuclei. Green asterisks points to the follicle sacs in which the sperm cells are stored. (b) Average total selenium measurement of male whole-body tissues from the non-Se supplemented and the Se-supplemented groups

Individual Oyster Tissue	Non-supplemented			Supplemented		
	Av. Se per individual (µg/g)	**SD**	**RSD**	**Av. Se per individual (µg/g)**	**SD**	**RSD**
A	3.47	0.27	7.50	80.2	0.80	1.10
B	3.3	0.20	6.15	48.4	0.34	0.69
C	3.12	0.25	7.95	179	0.28	0.16
D	4.72	0.23	4.94	36.8	0.43	0.17
E	2.9	0.18	6.10	19.9	0.32	1.59
F				52.1	0.52	0.98
G				13	0.20	1.56
Average Se concentration per group (µg/g of tissue)	**3.502**			**61.3**		
Certified Reference Materials NRCC, National Research Council of Canada NRCC Reference Material Site	***Av. Se µg/g***	***SD***	***RSD***			
DORM-1*	*1.46*	*0.03*	*2.27*			
TORT-2 **	*5.64*	*0.09*	*1.68*			

**certified value: 1.62 ± 0.12 µg/g (dogfish muscle)*

***certified value: 5.63 ± 0.6à µg/g (Lobster hepatopancreas)*

Table 4.3.2: Selenium Determination of individual male oyster tissue.
Total Se levels of individual male whole body soft tissues were determined by ICP-MS. Individual Se values from the non-supplemented and supplemented group are given. Standard Deviation (SD) and Relative Standard Deviation (RSD) was measured for each sample. Values for certified reference materials are indicated.

Material	Non-supplemented		Supplemented	
	Av. Se. µg/g	**SD**	**Av. Se. µg/g**	**SD**
Microalgae Food Source	0.0263	0.0015	0.7451	0.0063
Artificial Sea water (after trial)	0.0086	0.0016	0.1796	0.0021
Tissue Protein Extract	0.1511	0.0050	1.3567	0.0107
Tissue debris	0.2907	0.0107	0.5558	0.0021

Table 4.3.3: Selenium Determination of relevant materials from the supplementation trial.
Total selenium levels were determined for each sample indicated from the appropriate supplementation group. Tissue protein extracts are cytoplasmic proteins isolated after tissue lysis and the remaining debris were separated by centrifugation and sent for Se determination. Standard deviation values (SD) were also determined.

4.3.3 Selenium effects on selenoproteome expression in C.gigas

Whole body tissues of two individual animals from each supplementation group with similar levels of determined total selenium were pooled and prepared for ribosome profiling (ribo-seq) library generation and parallel RNA-sequencing (RNA-seq) (Table 4.3.4). Ribosome profiling allows deep sequencing and mapping of mRNA protected fragments along selenoprotein mRNA, revealing insights on translation at specific regions within an mRNA. After library preparation (Suppl. Fig 4.5.1 a), bio-informatic filters were applied to the raw sequencing reads (Suppl. Fig. 4.5.1 b) followed by sequence alignment to the full oyster transcriptome. Metagene analysis performed as a quality control step shows that the ribosome profiling samples prepared from each group exhibited characteristics of ribosome foot prints: there is correct abundance of ribosome foot-prints relative to initiation and termination positions at P and A-site positions of the ribosomes, some good triplet phasing and low 3' and 5' untranslated region coverage relative to annotated open-reading frames (Fig.4.3.3).

Supplementation Group	**Individual Se concentration 1 (µg/g)**	**Individual Se concentration 2 (µg/g)**	**Average Se conc. (µg/g)**
Non-Supplemented	3.47	3.3	**3.4**
Supplemented	13.3	19.9	**17**

Table 4.3.4: Selenium content of samples pooled for ribosome profiling and RNA-sequencing
Individual Se concentration and average selenium concentration are indicated for tissue samples used in cDNA library preparation.

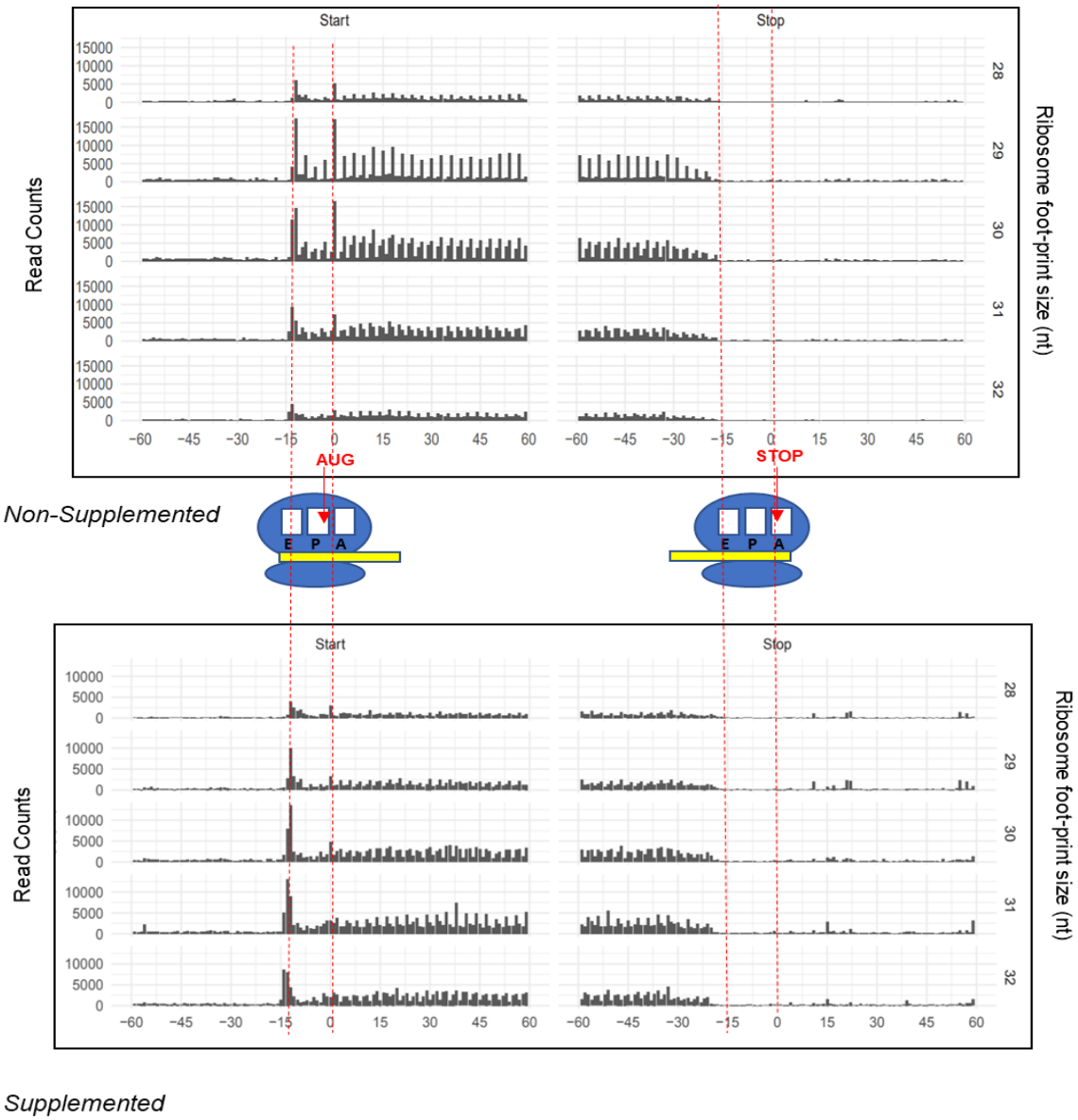


Figure 4.3.3: Metagene analysis of ribosome profiling libraries for quality control.
Pre-processed ribosome profiling reads from the non-supplemented and supplemented groups were aligned to an assembled *C.gigas* transcriptome. Ribosome protected fragments for each read-length (28, 29, 30, 31 and 32 nt) were analysed for triplet phasing and relative positions of the reads to the start, stop codons, coding regions and untranslated regions. Ribosome alignments denote that reads start abruptly at -12 position when AUG is at the peptidyl-site of the ribosome and abruptly ends at -15 when stop codon is at the amino-acyl site of the ribosome.

The rich data-set obtained from deep-sequencing alignments of ribo-seq and RNA-seq libraries allowed for qualitative comparison of the oyster selenoproteins and Sec-redefinition components expression in elevated selenium levels. For RNA-seq, only the CDS was used to obtain read counts as issues were detected in the assembled transcript including that of two individual transcripts being fused together in one annotation. For ribo-seq, RPF reads 15 nts from the annotated start and stop codons were not considered to avoid biased read counts as a result of ribosome accumulation at these positions from cycloheximide treatment. RPFs and RNA counts were then normalized against transcript length by expressing the values as reads per kilobase (kb) length of transcript per million mapped reads (RPKM or RPFKM). Transcripts with read-counts less than 15 were discarded to avoid mis-interpretation of Se-effects due to low coverage (AhPC.1, Di.2, SelenoH and psTK.1).

A general trend of increasing mRNA abundance (Fig 4.3.4b) and translation (Fig 4.3.4a) of selenoproteins was observed consistent with previous reports of ribosome profiling analysis after selenium supplementation in mice (Howard et al., 2013, Tsuji et al., 2015). Translation efficiency, which measures RPF abundance relative to mRNA abundance is shown to decrease with supplementation (Fig 4.3.4 c) due to a stronger increase in mRNA levels, with individual selenoprotein mRNAs demonstrating around a 7-fold increase, compared to RPFs, demonstrating up to about a 3-fold increase in the supplemented samples (Fig 4.3.5). The observations in oyster differ from that reported in mice where the Se-effect is greater at the protein rather than at transcript level. This discrepancy could be attributed to a differential rate of mRNA turn-over, stability or susceptibility to non-sense mediated decay between organisms. In addition, minimal Se regulation was detected for the components of Sec-insertion machinery (eEFSec, SecS and SBP2) (Fig 4.3.5).

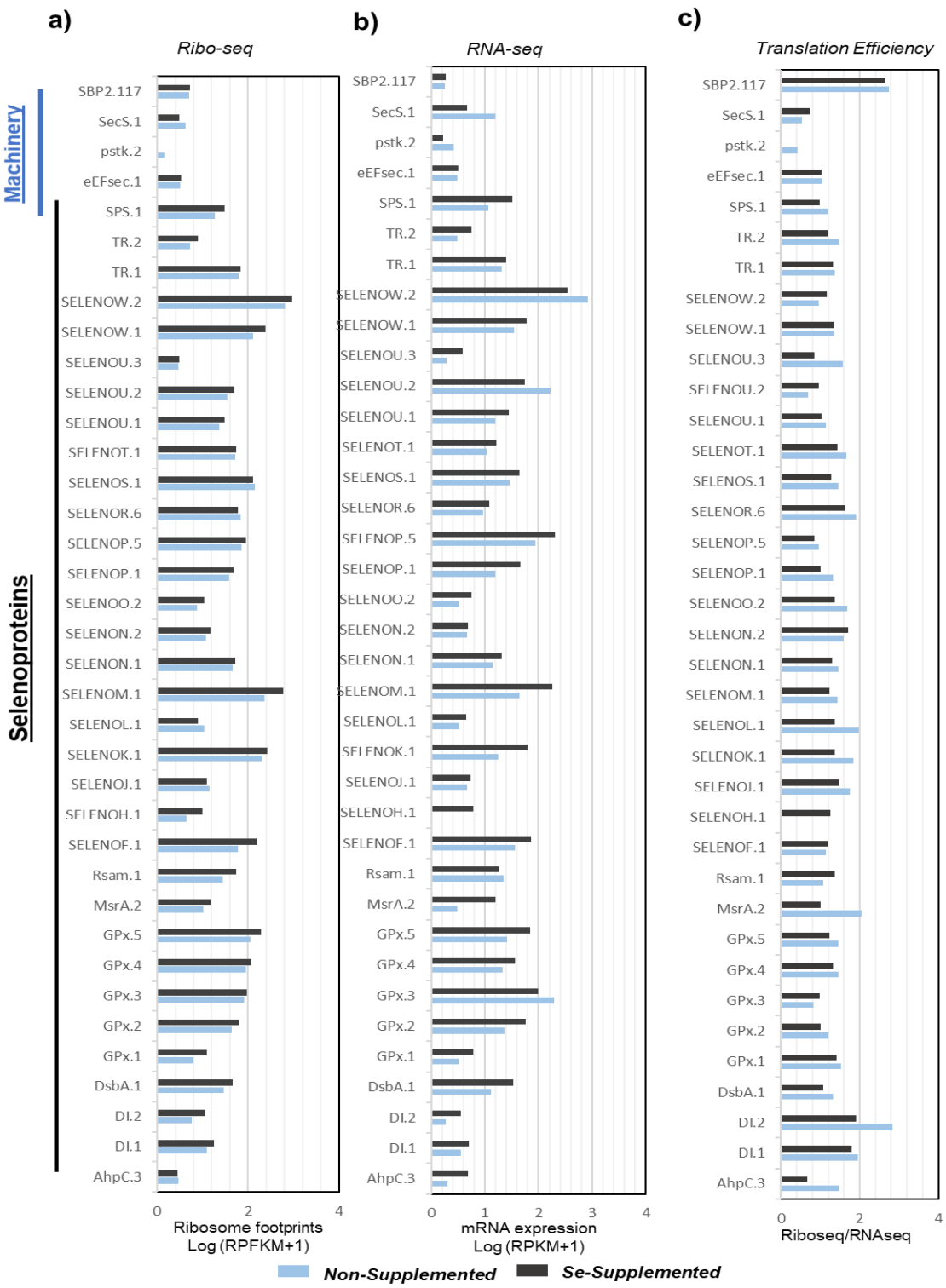


Figure 4.3.4: Qualitative effects of selenium on the expression of selenoproteins and selenoprotein machinery.

a) Ribosome foot-prints from ribosome profiling and b) mRNA expression from RNA-sequencing normalized as RPKMs/RPFKMs (RNA reads or RPF reads per kb length of transcript per million mapped reads) and expressed as log (RPKM+1) for comparison for each selenoprotein and selenoprotein machinery component. c) Translation efficiency is expressed as a ratio of ribosome footprints to mRNA abundance (RPFKM/RPKM).

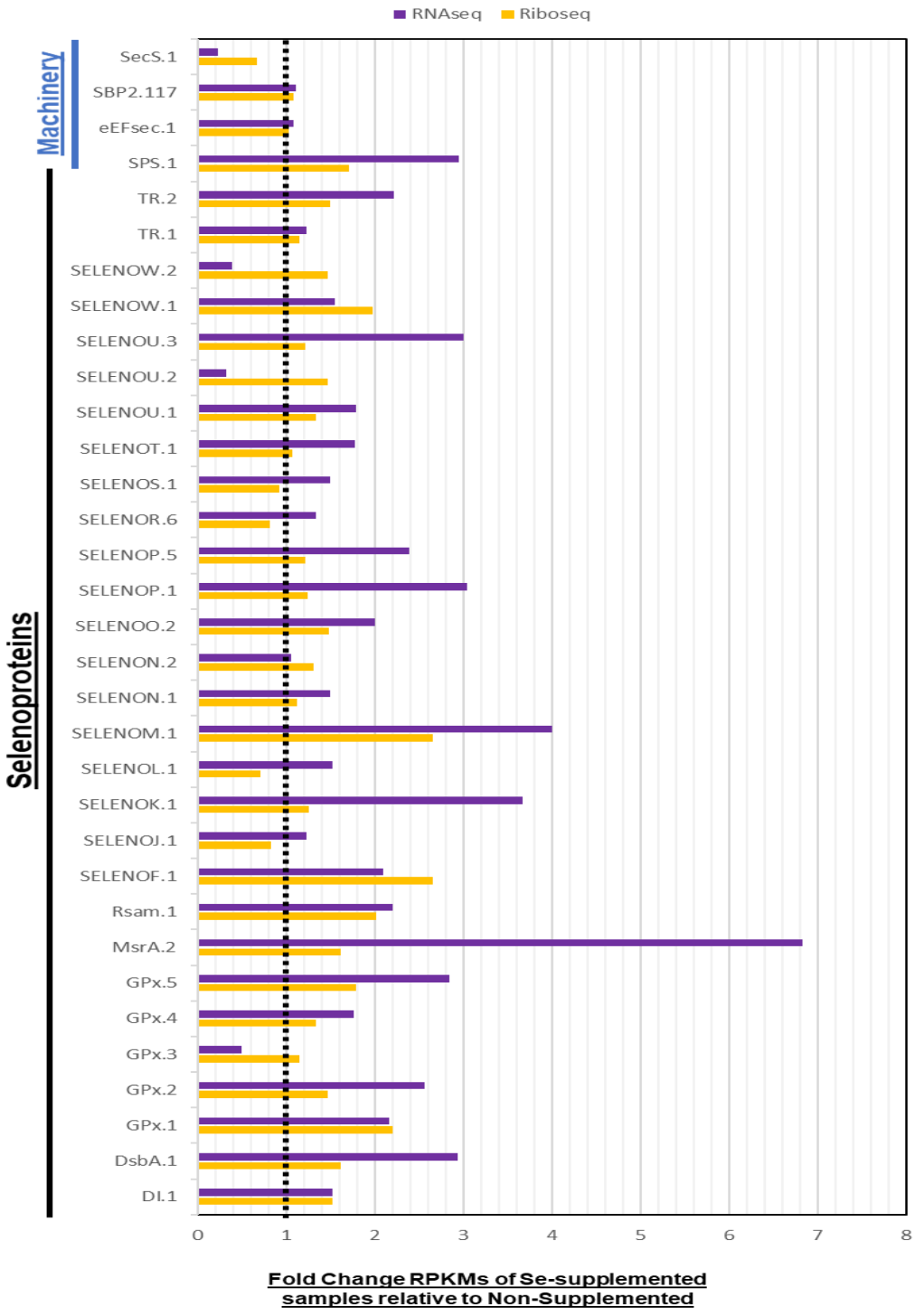


Figure 4.3.5: Comparison of selenium effects on mRNA-abundance (RNA-seq) and translation (Ribo-seq) for each selenoprotein and machinery.

mRNA abundance (purple) and ribosome footprints (yellow) are expressed as fold-change of the RPKMs of genes from the Se-supplemented samples relative to the non-supplemented sample which is given the value of 1 (depicted by the black dotted line).

4.3.4 Effects of Selenium on UGA-Redefinition Efficiency (URE) and selenoprotein translation regulation

Previously proposed methods (described in section 4.2.13, B) (Howard et al., 2013, Fradejas-Villar et al., 2017) were used to obtain a surrogate measure of Sec-UGA redefinition efficiency (URE). URE is expressed as a percentage of 3'TE to 5'TE to represent the percentage of ribosomes that translated past the Sec-UGA codon compared to those that initiated translation (Fig 4.3.6 a). URE could not be determined for selenoprotein genes that have a Sec-UGA codon proximal to the annotated stop-codon (Rsam.1, SelenoK, SelenoO, SelenoT, SelenoS) and those with Sec-UGA 5' of the selenoprotein gene sequence (Dsba, MsRa, SelenoW). URE calculations were also omitted for SelenoL as it was found to have 2 Sec codons adjacent to each other. SelenoP UGA redefinition efficiency was determined by calculating 5'RPFs up to UGA 1 and 3' RPFs from UGA 1 to UGA 2. These reads were normalized with mRNA abundance at the CDS to obtain 5' and 3' translation efficiency.

We observed that, consistent with previous reports of Sec-UGA redefinition in mice (Howard et al., 2013), most oyster selenoproteins exhibited an inefficient URE with most rates below 10%, even after supplementation (Table 4.3.5). Exceptions to these finding are Di.1, Gpx.4, SelenoF, SelenoJ and SelenoN which demonstrated efficient Sec-UGA redefinition rates of more than 50%. It is important to note that a known artefact of ribosome profiling is RPF accumulation at the initiation sites which occurs as a result of blocking elongation. This was circumvented by excluding reads 15 nt from the start codon. Excluding these reads for genes that do not demonstrate this artefact at initiation may skew estimated URE calculation rates.

In general, Se-supplementation exhibited differential 5' and 3'RPF regulation. Not all selenoproteins exhibited an increase in URE or any strong 3'RPF increase with supplementation contrary to mice findings where a strong 3'RPF increase was observed (Howard et al., 2013). For mammalian SelenoP UGAs, selenium level effects predominantly influence UGA1 redefinition efficiency (Howard et al., 2013; Fradejas et al., 2017). For oyster SelenoP, we compared normalized read numbers for all the CDS 5' of UGA1 with those from the region between UGA1 and UGA2. The URE for non-supplemented group was 4.68% whereas that for the supplement group was 6.8% (Table 4.3.5, SelenoP.5 b). A strong RPF peak mapping to CDS positions

1-20 from AUG was observed. Repeating the calculation excluding the pause site, which could be an alternative regulatory event independent of UGA1, yielded URE values of 5.5% and 4.7% for non-supplemented and supplemented, respectively (Table 4.3.5, SelenoP.5 a). Unlike its mammalian counterpart, elevated selenium level does not directly influence UGA1 redefinition. The role of the pause site is considered in the Discussion.

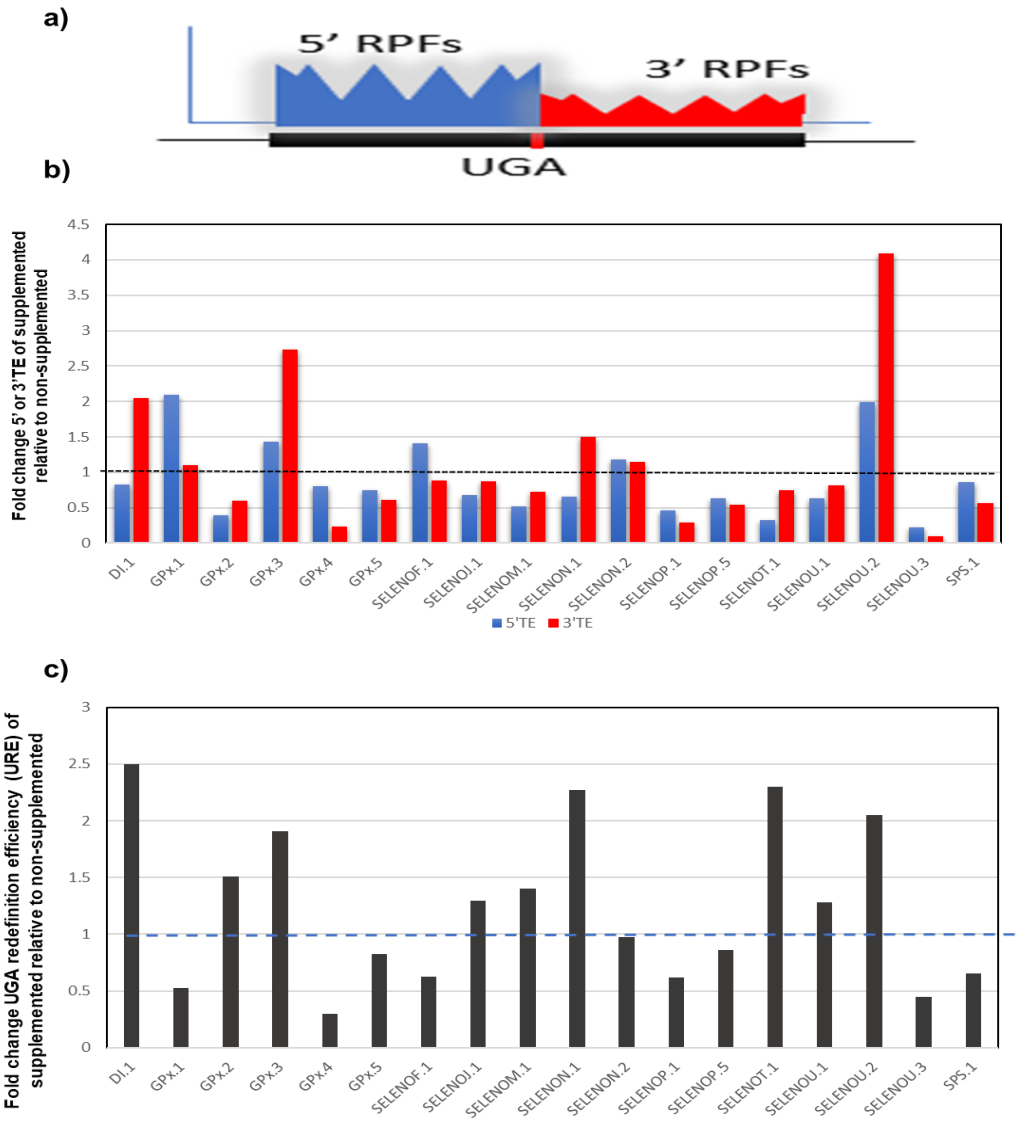


Figure 4.3.6 Analysis of selenoprotein translation 5' and 3' of Sec-UGA.
a) Schematics of RPF reads obtained 5' (5'RPF) in blue and 3' (3'RPF) in red relative to the Sec-UGA codon. b) 5'RPFs (blue) 3' RPFs (red) for each selenoprotein gene of the supplemented samples. These were normalized to RNA reads (RPKM) per gene expressed as fold change relative to the non-supplemented samples given the value of 1 (black dotted line). c) UGA redefinition efficiency (URE) per selenoprotein gene. URE for each gene were calculated by expressing translation efficiency at 3' of Sec-UGA (3'RPF/RNA) as a ratio of translation efficiency at the 5' of Sec-UGA (5'RPF/RNA). The selenium effects to URE of each selenoprotein was expressed as fold-change relative to non-supplemented samples valued as 1 (blue line).

	Non-Supplemented			Se-Supplemented		
SELENOPROTEINS	**5'TE**	**3'TE**	**% URE**	**5'TE**	**3'TE**	**% URE**
DI.1	6.865114	2.733514	39.81745	5.641261	5.606082	99.3764
GPx.1	33.74046	0.956093	2.83367	70.80081	1.053083	1.487389
GPx.2	21.817	1.347086	6.174477	8.608656	0.800278	9.296206
GPx.3	16.65069	0.360296	2.163853	23.84878	0.985089	4.130562
GPx.4	8.794694	17.83616	202.806	7.07051	4.236248	59.91432
GPx.5	129.9663	5.681988	4.371894	96.79251	3.486436	3.601969
SELENOF.1	4.605308	5.38931	117.0239	6.469076	4.734556	73.18752
SELENOJ.1	4.332765	7.054533	162.8183	2.936342	6.188029	210.7394
SELENOM.1	128.4129	7.649625	5.957056	66.1427	5.522985	8.350106
SELENON.1	4.079624	2.211224	54.20167	2.690665	3.315987	123.2404
SELENON.2	3.250611	3.085206	94.91157	3.853749	3.55281	92.191
SELENOP.1	109.5829	2.540594	2.318422	51.12642	0.73374	1.435148
SELENOP.5 (a)	**19.67976**	**1.093151**	**5.554694**	**12.42015**	**0.593207**	**4.77617**
SELENOP.5 (b)	**10.6**	**0.49**	**4.6**	**23.1**	**1.59**	**6.8**
SELENOT.1	66.32628	3.865746	5.828377	21.40883	2.868419	13.3983
SELENOU.1	7.150711	1.528405	21.37416	4.550539	1.246615	27.3949
SELENOU.2	4.962322	0.177186	3.570635	9.883407	0.724633	7.331818
SELENOU.3	5.796359	1.013572	17.48636	1.317667	0.102385	7.77021
SPS.1	86.3302	1.317051	1.525597	74.11442	0.743365	1.002997

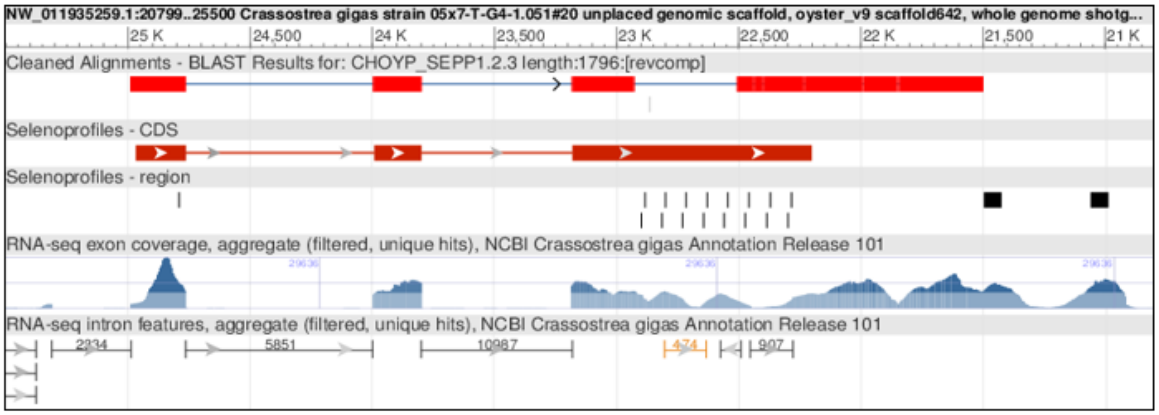
Table 4.3.5: Percentage UGA redefinition efficiency (URE) calculations per selenoprotein.
Selenoprotein gene UGA redefinition efficiency was obtained by determining 5'TE (5' RPFs measured from CDS excluding 15 nt from AUG, up to Sec-UGA 1 for SelenoP and normalized to RNA RPKMs) and 3' TE (3'RPFs from Sec-UGA to annotated stop or Sec-UGA 2 for SelenoP and normalized to RNA RPKMs). URE is the ratio of 3'TE to 5'TE expressed as a percentage from each supplementation group.

4.3.5 Gene duplication of SelenoP in C.gigas

Selenoprotein predictions in *C.gigas* revealed two *selenoP* genes which contains differing numbers of multiple in-frame Sec-UGA. These *selenoP* sequences were identified both in the genome and the transciptome. One gene, referred to here as *selenoP2*, has 17 UGA codons (Fig 4.3.7 a) and the second gene, referred here as *selenoP1/selenoP*, has 46 UGA codons (Fig 4.3.7 b). The number of exons and intron boundaries are conserved between the two selenoP genes.

Alignment between the two transcript sequences revealed a highly conserved 5' half (Fig 4.3.8 a1). The sequence identity in the 3'UTR on the other hand, is significantly lower between the two genes apart from the SECIS elements which appeared to be conserved (Fig 4.3.8 a4). SECIS 1 is almost identical between the two genes while SECIS 2 alignments shows a few substitutions predominantly in stem 1 (Fig 4.3.8b). Interestingly, part of the CDS in the Sec-rich region of *selenoP1* has homology with the UTR of selenoP2 (Fig 4.3.8 a3), suggesting that a region of *selenoP1* CDS became UTR in *selenoP2* after duplication. Further analysis of the *selenoP2* at the protein level revealed a very distinct repetitive amino-acid motif highlighted in yellow and black in Fig 4.3.8 c. These regions consist of eight instances of repeats containing two Sec-UGAs with a few amino-acid substitutions. Those two UGAs correspond to UGA2 and UGA3 in *selenoP1* (Fig 4.3.8 a2). A second selenoP gene (selenoPB) has been previously reported to occur in vertebrates (except placental mammals) containing a single Sec-UGA and one SECIS element (Mariotti et al., 2012). A brief comparison of transcript expression of the two SelenoP mRNA revealed a much higher expression of selenoP1 with 46 Sec-UGA compared to selenoP2 with 17 Sec-UGA (Fig 4.3.7a and b).

a)



b)

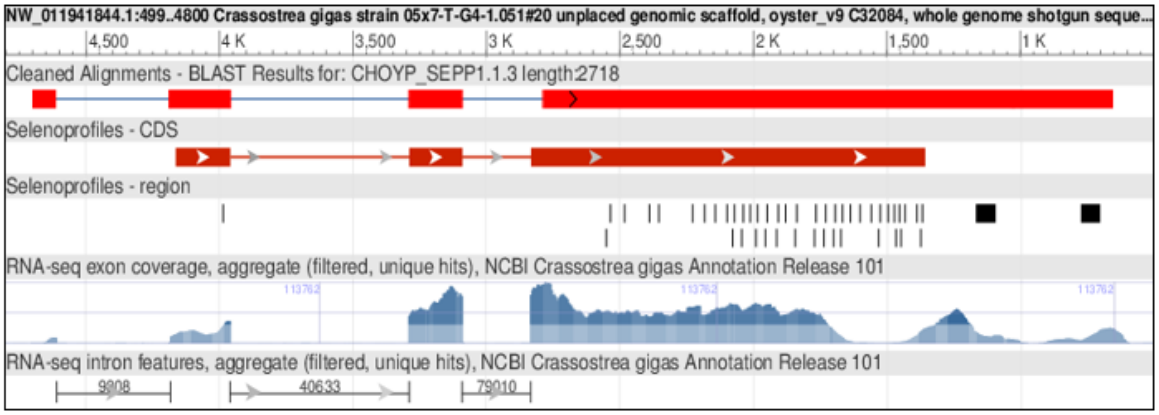


Figure 4.3.7: Gene structure a)17-Sec-UGA SelenoP2 and b) 46-Sec-UGA selenoP1 in *C.gigas*. The top track (light red) corresponds to the transcript mapped with BLAST onto the genome. The second track (dark red) corresponds to CDS. The third track indicates the position of in-frame UGAs (thin bars) and the two SECIS (wide bars). The fourth track corresponds to aggregate RNA-seq coverage. Bottom track corresponds to intron boundaries. The last two tracks were provided by NCBI. Visualized using the NCBI Genome Data Viewer (GDV).

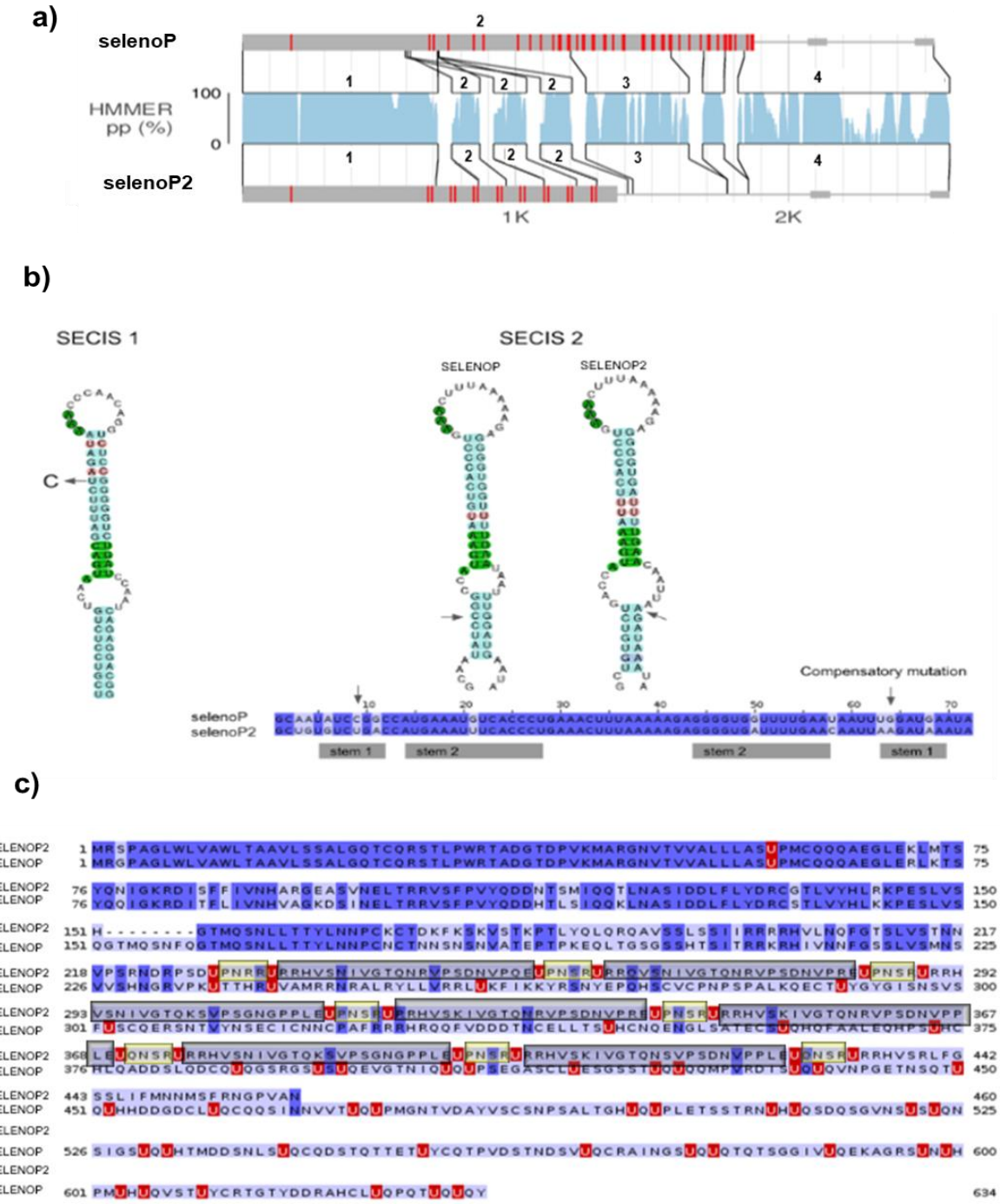


Figure 4.3.8: Characterization of selenoP2 with 17-Sec-UGA.

a) Transcript sequence alignment of the two selenoP genes. b) Sequence and secondary structure of SECIS1 (left) and SECIS2 (right) from the two selenoP genes in *C.gigas*. Nucleotide sequence alignment for SECIS 2 is shown. The arrow points at compensatory nucleotide mutation in the SECIS elements of SelenoP2. c) Protein sequence alignment of the two SELENOP in C.gigas. SELENOP2 is with 17-Secs (U) and SELENOP is with 46-Secs (U) and are highlighted in red. Conserved amino acids and positions between the two SELENOP are highlighted in dark-blue. The repeat amino-acid sequence motif in SELENOP1 is highlighted in yellow and black.

Using our generated ribosome-profiling and RNA-seq data sets, we confirm low levels of transcription and translation of *selenoP2* with 17-Sec-UGA. A coverage map was generated (Fig 4.3.9) which determined RNA-reads up to the end of selenoP2 transcript confirming transcription. We also observed RPFs with high abundance mapping to the N-terminal region of *selenoP1* before UGA 1 followed by a significant decrease in RPFs in the Sec-rich region. Alignment of RPF reads in this case poses issues including that of a very high sequence homology in the N-terminal domain of the two selenoP genes which might infer reads mapping to either of the two genes in this region.

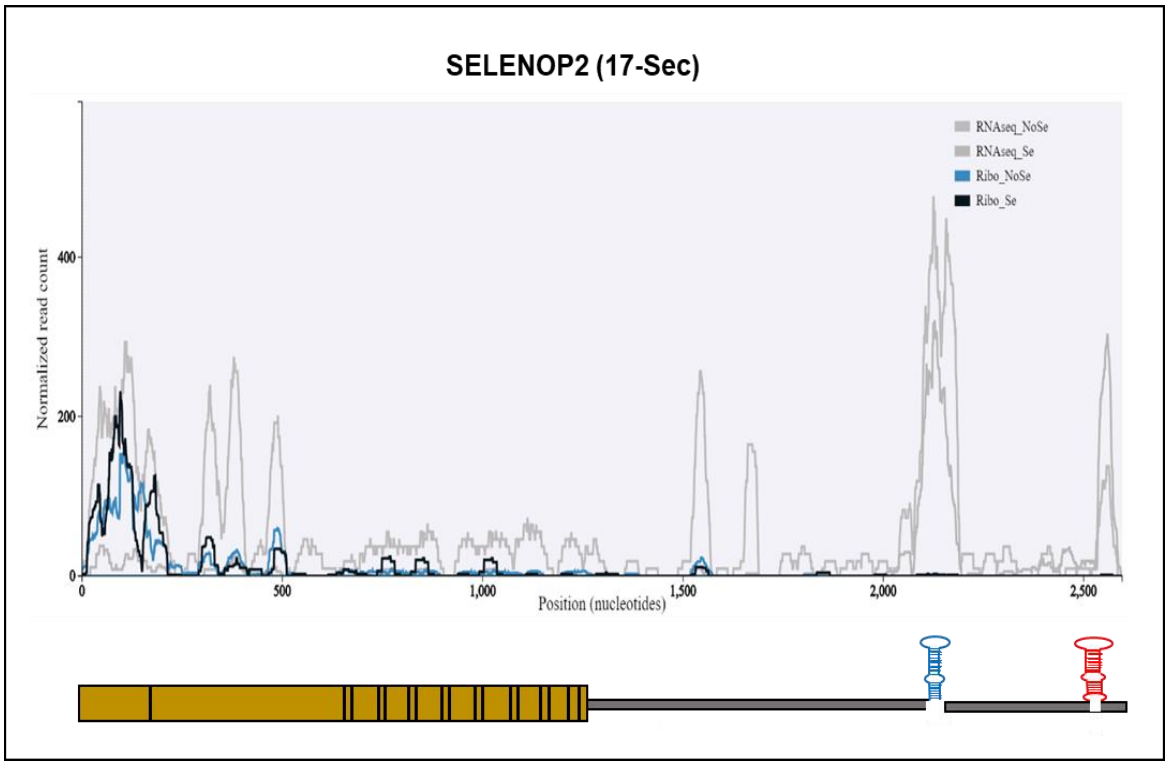


Figure 4.3.9: RPF and mRNA coverage map for SELENOP2
Blue and black plots are from ribosome profiling while the grey plots are from RNA-sequencing. A transcript model for SELENOP2 is shown where the reads are aligned. The grey box denotes coding region, green lines for Sec residues, purple structures are SECIS 1 and SECIS 2 respectively and thin grey line is 3'UTR. The y-axis represents normalized read counts for comparison while the x-axis shows nucleotide position.

4.3.6 Monitoring in-vivo SELENOP translation regulation

It was previously shown that the reconstituted rabbit reticulocyte lysate (RRL) system can only support production of oyster SELENOP termination products (Chapter 3, Section 3.3.2) suggesting insensitivity of the vertebrate translation components to signals in oyster selenoP mRNA and/or unknown oyster trans-acting components lacking in RRL, necessary for processive Sec incorporation. This warranted investigation of SelenoP translation *in-vivo*.

To confirm translation of oyster SelenoP with 46 Sec-UGA, RPFs and RNA-seq reads were mapped specifically to SelenoP mRNA allowing for ambiguously mapping reads due to the presence of the second SelenoP2 gene in the annotated transcriptome. RNA-seq showed full coverage along oyster selenoP mRNA, including its 5' 110 nt Leader, CDS and 3' UTR (Fig 4.3.10). Notably, some RPFs mapped to a specific 61-nt region spanning the 5'UTR. Upon sequence analysis, a 16 codon Leader ORF was identified in a modest Kozak consensus c.G(AUG)AA. Ribosome coverage was only moderately greater in the sample derived from selenium supplemented tissue. In the non-supplemented sample, the most abundant fragments are mapped from 15 nts 5' of the CDS start and sharply drop 20 nts 3' of it (Fig 4.3.10, green arrow). The ISL RNA structure is found 20 nt 3' of the coverage drop (ISL at positions 37-157 from main ORF AUG). The amount of these fragments is greatly reduced upon selenium supplementation.

In the supplemented material there are also abundant fragments approximately from nts 17- 35, 38-128 and 149-166 3' of the main ORF. In this region the fragments from the non-supplemented material are somewhat less abundant and some of their boundaries moderately shifted. Inferences from such complex patterns in profiling data merit caution, but the data show a broad RPF accumulation prior to UGA1 that is also present in mammalian SelenoP ribosome profiling data (Fradejas-Villar et al., 2017, Mariotti et al., 2017). As introduced above, a second gene, termed oyster SelenoP2 was identified in *C.gigas* exhibiting high homology to SelenoP, 5' of UGA1. However, we detected very minimal RPF coverage 3' of UGA1 in SelenoP2 where the sequences are less homologous, suggesting that its expression levels are significantly lower, if not almost undetectable, compared to SelenoP. We therefore deduce that reads obtained in the coverage map should accurately represent SelenoP (with 46 Sec-UGA) translation.

The oyster RPFs map to the full length of the predicted CDS up to the UAG termination codon (Fig 4.3.10). This result strongly supports full-length 46 Sec-UGA SelenoP translation in *C. gigas*. Translation of UGAs 2-46 appears to be continuous and processive as indicated by the absence of ribosomal pausing and approximately equal RPF coverage across the UGA-rich segment, comparable to that found for its mammalian counterpart translation, both *in vitro* (Fixsen and Howard, 2010) and in mice (Howard et al., 2013).

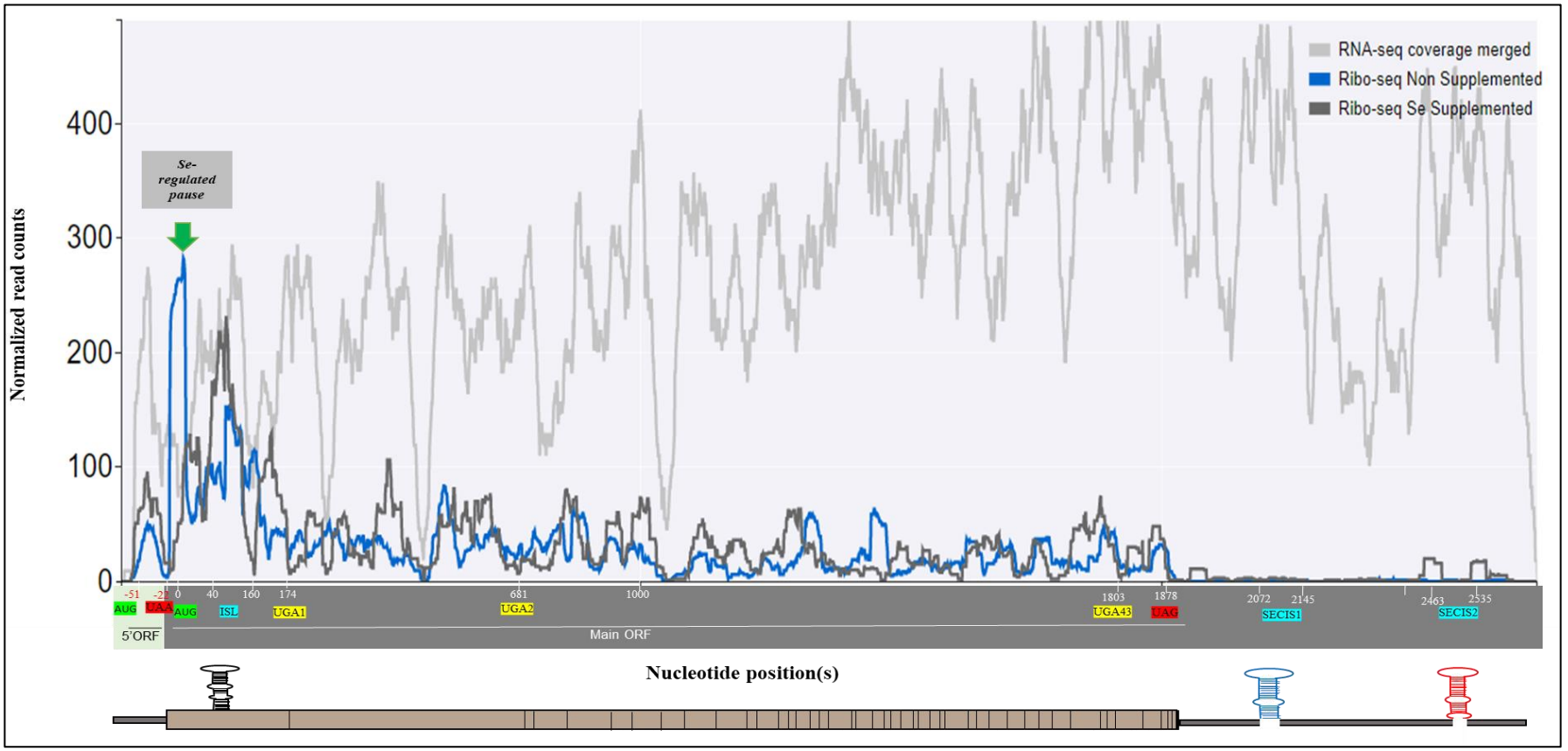


Figure 4.3.10: Monitoring SelenoP translation by ribosome profiling

RNA-seq reads (light-grey) and riboseq RPFs (blue for non-supplemented sample, dark grey for selenium supplemented) aligned to SelenoP mRNA positions. The RNA-seq reads were merged for the supplemented and non-supplemented samples to represent RNA coverage across the full transcript. Two open reading frames (ORF) were identified: the 5' leader ORF spans -55 to -22 nt positions (light green) whereas the start of the main ORF position is depicted as position 0 (grey). The boundaries of RNA structures identified in the main ORF are indicated. The green arrow indicates a selenium-regulated pause at initiation.

We confirmed our ribosome profiling results by oyster SELENOP detection using immunoblots of an aliquot from the same lysates used in ribosome profiling experiments with custom antibodies designed for SELENOP. The antibodies target the N-terminal region (anti NT SELENOP) before UGA 1 and the C-terminal region (anti-CT SELENOP) after UGA 43 (Fig 4.5.11a). These antibodies were tested for peptide affinity and specificity confirmed by antibody recognition of a tagged protein containing their immunogenic peptide sequence (Suppl. Fig 4.5.3 b and c). Specificity is better demonstrated by the N-terminal antibody compared to the C-terminal antibody. C-terminal antibody was used in parallel to N-terminal detection, in order to confirm the presence of the long forms/ full-length SELENOP.

The presence of an immunoreactive protein that runs at an apparent molecular weight of 68 kDa which matches the predicted molecular weight of full-length SELENOP (Fig 4.3.10 b and c, red arrow), recognized by the N and C-terminal antibodies was detected. Consistent with RPF abundance determined for SELENOP where RPFs are only modestly increased (Fig 4.3.4 a), the band intensity appeared to be slightly upregulated between the Se-supplemented and non-supplemented samples (1.1 fold higher in the N-terminal (Fig 4.3.11 b) and 0.3 (Fig 4.3.11 c) fold higher in the C-terminal). The C-terminal antibody recognized an additional pronounced band that runs below 37 kDa. Since this is not recognized by the N-terminal antibody, it is unlikely to be a shorter form of the protein terminating at the earlier UGAs. Proteolytic cleavage and post-translational modification of SELENOP could be a probable explanation, although antibody characterization (Suppl. Fig 4.5.2c) suggests that C-terminal antibody may recognize non-specific proteins.

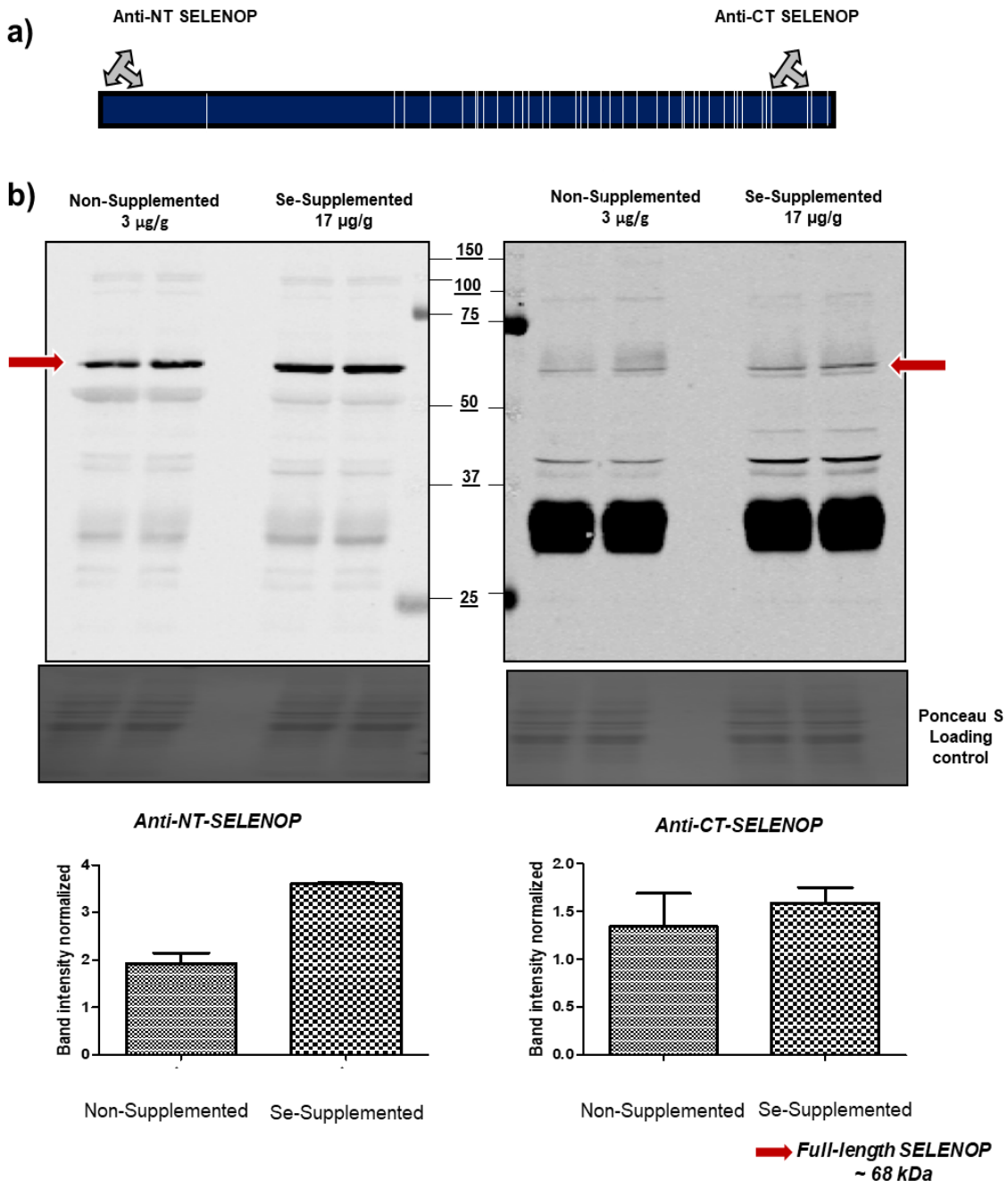


Figure 4.3.11: Immunoblot detection of SELENOP.

(a)Protein map of SELENOP indicating immunogenic targets of N and C-terminal antibodies. (b and c) Immunoblot of the tissue lysates used for library preparation using SELENOP custom antibodies targeting (b) N-terminal and (c) C-terminal portions of the protein. Red arrow indicates full-length protein detected at around 68 kDa. A densitometry graph (n=2 per sample) depicts the change in FL band intensity in the non-supplemented vs Se-supplemented sample from each antibody where the bands are normalized against Ponceau-S loading control.

4.3.7 SELENOP Sec-UGA incorporation

Human SelenoP contains 10 in-frame UGA codons, and they can also specify Cys instead of Sec to some degree depending on the selenium supply level (Turanov et al., 2015, Turanov et al., 2011). Since our ribosome profiling data strongly suggests that full-length SelenoP is translated including all 46 Sec-UGA, we wanted to confirm whether Sec is incorporated in SELENOP *in-vivo*. Radioactive ^{75}Se labelling has been the hallmark technique for studying selenoprotein expression as selenium is incorporated as the 21st amino-acid selenocysteine (Yim et al., 2018). Parallel western blot probing using antibodies specific to selenoprotein of interest helps to determine the identity of labelled proteins produced. Using published genomic and transcriptomic data on *C.gigas*, we determined transcript expression of SelenoP at different developmental stages and conditions (Zhang et al., 2012) and confirmed expression of full-length mature transcript from the oyster larval stages (Fig. 4.3.12 a and b).

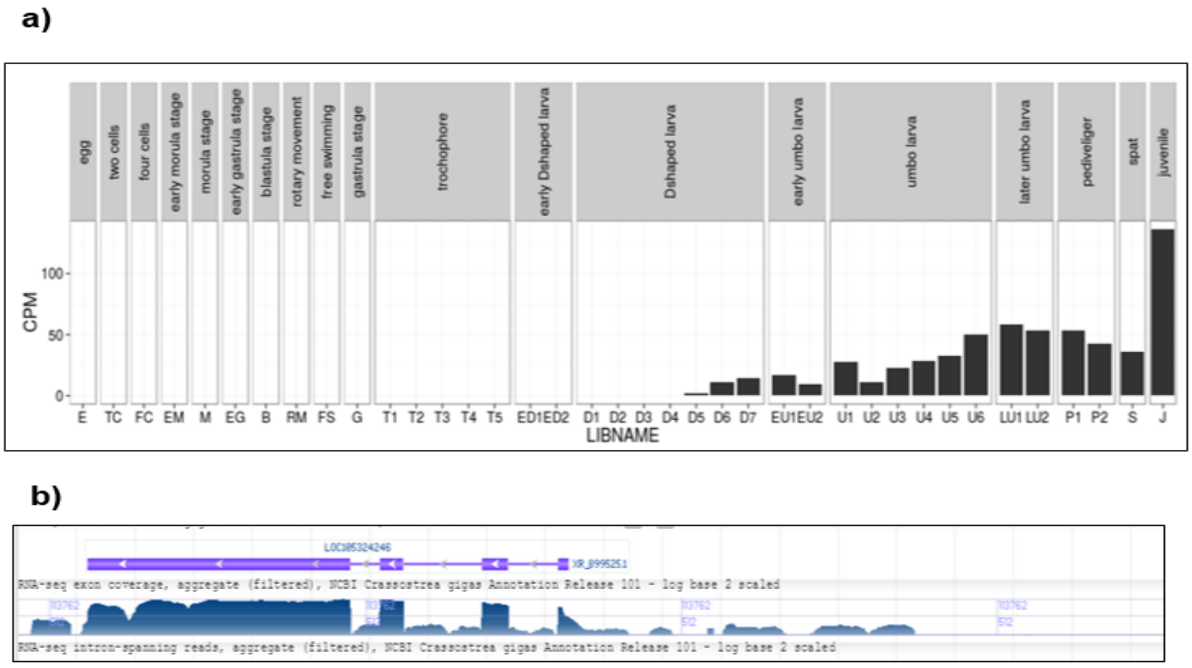


Figure 4.3.12: SelenoP transcript expression of oysters.
(a) Analysis of transcript expression for selenoprotein P at various stages of development from egg to juvenile stage obtained from Zhang *et al.*,2012 paper. CPM: counts per million reads, LIBNAME: library name for each stage indicated in the headings. (b) mRNA reads for SelenoP mapping to the full-length of the transcript at the larval stages. SelenoP gene contig (purple) indicates exon (thick line) and intron boundaries (thin line). The reads across SelenoP region is indicated by the blue aggregate.

Free-swimming larvae at 7-day and 14-day old stages were gifts from an oyster cultivation facility in Washington, USA (*Pacific Sea Foods Quilcene WA 98376*). Oyster larvae (100-200) at each respective stage were labelled by the addition of 100 µM ^{75}Se diluted in artificial sea water and algae food source to promote filter feeding. Labelled larvae were analysed for SELENOP production by autoradiography and a parallel Western blot was performed revealing radiolabelled selenoproteins in *C.gigas* (Fig.4.3.13 a). A protein product of 68 kDa was detected by both the SELENOP N and C-terminal antibodies (Fig 4.3.13 b and c), strongly suggesting Sec-incorporation in SELENOP *in-vivo*. The presence of lower molecular weight bands detected by autoradiography as well as by the N-terminal antibody (Fig 4.3.10 b) suggests a possible termination product, while the other radio-labelled products not detected by the antibody might correspond to other, more abundant selenoproteins. C-terminal antibody recognition, albeit less apparent, suggests full-length Sec-incorporation in SELENOP validating ribosome-coverage from profiling data.

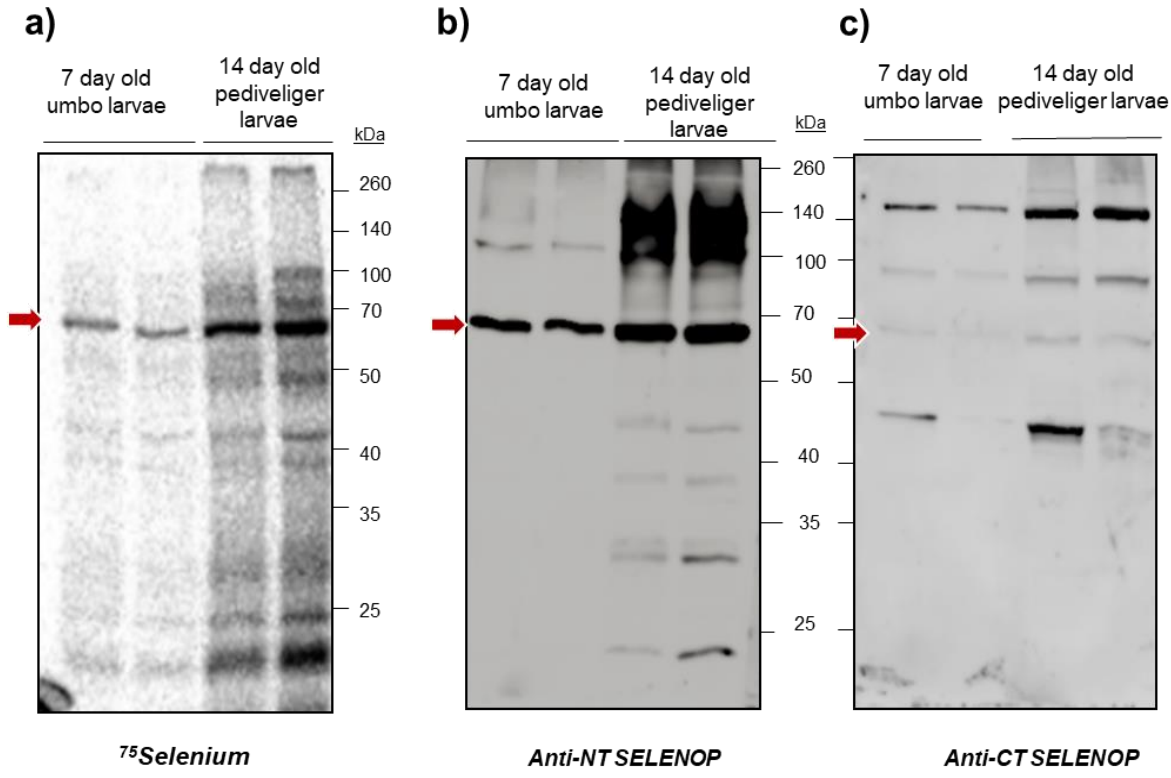


Figure 4.3.13: ^{75}Selenium labelled proteins in *C.gigas* larvae.
a) Autoradiography of radio-labelled proteins in oyster larvae. b) Anti-N-terminal SELENOP and c) Anti-C-terminal SELENOP immunoblots of oyster larvae proteins. Red arrows (FL) indicate a radio-labelled and immunoreactive bands at an apparent molecular weight of 68 kDa which is approximate to size predicted of full-length SELENOP.

4.4 Discussion

C.gigas selenoproteome, selenium uptake and utilization

In this chapter, we explored the oyster selenoproteome in detail and provided insights into Se effects on selenoprotein mRNA expression, translation and UGA redefinition efficiency *in-vivo*. We first confirmed expansion of the oyster selenoproteome by identifying 31 selenoproteins, greater than their mammalian counterparts, thus, consistent with reports of larger selenoproteomes identified in aquatic species compared to terrestrials (Lobanov et al., 2007). We provide the first report of a Sec-containing RSAM protein in oysters and have found that it is also present in other bivalves, cnidarian and crustacean species (Fig 4.3.1). The RSAM protein in echinoderms had replaced Sec with Cys suggesting that Sec/Cys interchangeability is dynamic as previously discussed in Chapter 3. In addition, a second SelenoP gene (SelenoP2) with a reduced number of Sec-codons (17-Sec UGA) in its C-terminal domain was identified in *C.gigas*. This gene is expressed significantly less both in the transcript and protein levels compared to SelenoP with 46-Sec UGA.

Supplementation studies revealed that the oysters were able to accumulate extremely high levels of Se in their tissues, up to about 50-fold per individual. Analysis of mRNA and protein levels in response to supplementation (samples used for library preparation had around 6-fold increase in total Se levels), demonstrated a much stronger response at the mRNA level and a modest upregulation of translation. This suggests that for the oysters, Se levels are directed more towards transcript production rather than protein synthesis. With relevance to this observation, it is also well known that mammalian selenoproteins are differentially regulated at transcription. For instance, Gpx1 is a known target of non-sense mediated decay (NMD) (Moriarty et al., 1998, Sunde and Raines, 2011) and thus its mRNA levels are highly affected by Se status. SelenoW has also been shown to have a high rate of mRNA turn-over (Gu et al., 2002). Such transcript levels increase in response to higher Se demonstrated by the oyster selenoproteins suggest that the mRNA stability is highly improved by Se-status possibly by promoting Sec-decoding and reducing NMD susceptibility. Further, we observed variable 5' and 3'RPF regulation in response to Se. Not all selenoproteins exhibited an increase in URE or any strong 3'RPF increase with supplementation, contrary to findings in mice where a very strong 3'RPF increase was observed in

response to a higher Se levels (Howard et al., 2013). The variability in expression observed in the oyster selenoprotein mRNA levels and Sec-UGA redefinition with Se-supplementation suggests possible regulatory differences between organisms.

RPF abundance was also upregulated with individual selenoproteins exhibiting up to a 3-fold increase in RPFs in response to Se. At least for SELENOP, RPF abundance correlated with protein abundance observed by immunodetection (Fig. 4.4.10 b and c). In addition, SELENOP probing of tissues with higher total Se-levels demonstrated that full-length SELENOP production did not exactly match the increase in total Se observed of up to 26-fold (Suppl. Fig 4.5.3). The modest upregulation of selenoprotein levels determined by the profiling experiment suggests that the increased total Se levels determined in their tissue cannot be accounted for by just protein production, contrary to mammalian Se uptake where most Se is utilized for Sec incorporation into selenoproteins (Burk and Hill, 2015). It is tempting to speculate that perhaps in oysters, a dedicated transcription factor which binds Se is present and thus might explain the stronger increase in mRNA levels observed. In addition, non-protein components that bind Se, for instance, Se attaching to sugars or other low molecular weight selenocompounds, could also be a possibility.

Our analysis highlights some interesting differences in the regulation of the oyster selenoproteome by dietary Se compared to its mammalian counterpart at both mRNA level and ribosome activity. One obvious variable is the different biological systems presented in mammals compared to that in the oysters. It has been previously reported that in the oysters, accumulated minerals are compartmentalized resulting in a higher turn-over rate. Further, subcellular compartmentalization of metals in bivalves may contribute to detoxification and may explain how they can circumvent toxicity as a result of bioaccumulation (Rainbow and Smith, 2010). Oysters also possess an open circulatory system compared to a closed system in mammals. Such biological differences raise the question of whether the selenoproteins identified in oysters exhibit the same function. For instance, SelenoP in mammals has been strongly implicated in Se transport where a hierarchy of distribution is shown by preferential uptake of long SELENOP isoforms to the brain and the testes (Hill et al., 2003, Burk and Hill, 2015). In oysters, we cannot exclude a possible role of SELENOP as a Se-storage protein due to the high Sec content in its C-terminal domain. Se and SELENOP have also been implicated in mercury chelation (Spiller, 2018). The high Sec content

of SELENOP could be a possible adaptation to the oyster's increased susceptibility to accumulate toxic metals as an implication of filter-feeding (Hedge et al., 2009).

Although RNA-seq and ribo-seq analysis provided some interesting insights on the oyster selenoproteome, our results should be interpreted as a qualitative trend rather than an absolute quantitative measure. More replicates and conditions are required to determine quantifiable Se effects on the oyster selenoproteome at protein and RNA levels. In addition, library construction biases may also influence read quantification at any given position. Nevertheless, follow-up experiments on these observations would illuminate our knowledge of Se biology and utilization in bivalves which has possible environmental and nutritional implications.

Insights to the translation regulation of oyster SelenoP in-vivo

Earlier attempts to deduce the mechanism of multiple Sec incorporation events *in-vitro* alluded to a model where Sec incorporation is inefficient since translation termination would compete with Sec-incorporation at any given UGA (Nasim et al., 2000). Thus, in the case of SelenoP, translation is expected to be energetically expensive for cells and would result in low protein yields. However, abundant long forms of SELENOP detected in rat plasma (Read et al., 1990) implied an efficient Sec-incorporation mechanism *in-vivo*. In oysters, the extreme expansion of UGA codons in SelenoP mRNA poses intriguing mechanistic decoding questions. Confirmation of oyster SelenoP translation was facilitated by ribosome profiling and revealed some unprecedented insights: RPF reads were obtained across the full-length of the transcript including all 46 UGAs and up to the annotated stop codon. Metabolic Sec-incorporation into oyster SelenoP was also established with ^{75}Se labelling of oyster larvae along with a parallel Western blot using antisera against SelenoP N and C terminal regions. In addition, translation prior to UGA 1 was shown to be a slow rate-limiting step with broad RPF accumulation observed upstream of UGA 1 (redefinition rate after UGA1 is less than 6%) followed by processive translation along UGAs 2-46 consistent with the model proposed by Fixsen and Howard (Fixsen and Howard, 2010).

We have demonstrated that *in-vitro* removal of the 5'leader sequence in SelenoP (5'UTR) reduced radioalabelled SELENOP product and ISL-context

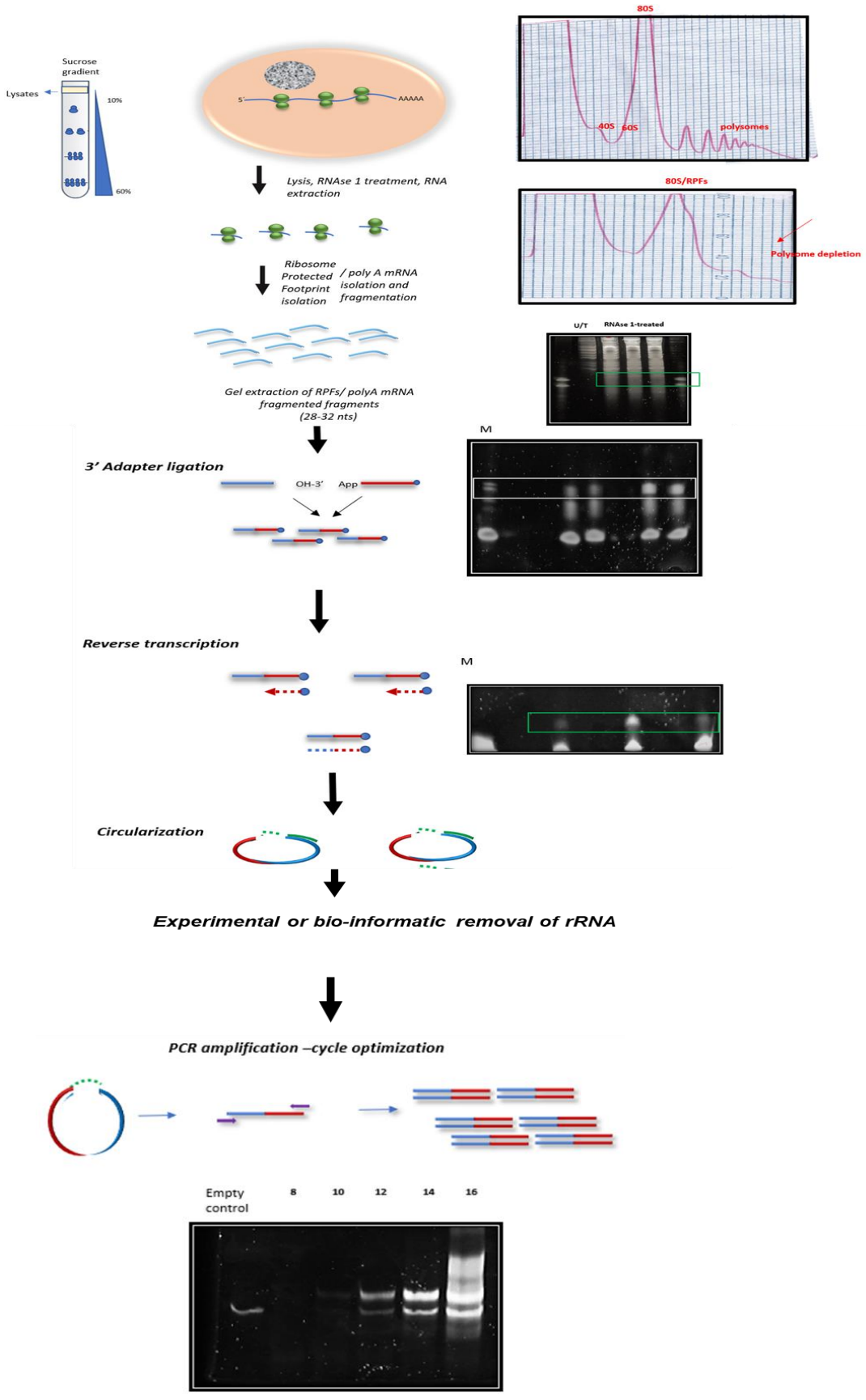
disruption resulted to the absence of ^{75}Se radiolabelled product, indicating its important role in SelenoP translation (Chapter 3). Our ribosome coverage map for oyster SelenoP revealed RPF fragments mapping to a 16-codon leader ORF in C.gigas SelenoP whose UAA terminator is 22 nts 5' of the main ORF AUG. The amount of this fragment is only slightly upregulated with selenium. In addition, a robust increase in ribosome occupancy surrounding -15 nt position up to +20 position from AUG, approximately corresponding to the size of a ribosome protected fragment during translation. This is suggestive of an alternative point of regulation at the site of initiation which is highly affected by Se status. There is a possibility of a strong ISL element effect occurring where translation initiation is regulated *in-vivo* at lower Se conditions. The RPF peak diminishes 20 nt 5' of the putative ISL element.

One possibility is that components of the pre-initiation complex, either RNA/protein, might interact with the ISL to allow for ribosome and initiation trans factors recruitment or that ISL would act as a 'starting block' to position the ribosome for an efficient translation initiation. Also, the absence of Kozak consensus (otherwise found in most oyster genes) in oyster selenoP may explain the requirement for a regulatory mechanism at initiation in this case. Reported cases in the *fepA* mRNA in bacteria identified a stem-loop-dependent 'starting block' mechanism to promote formation of the ternary initiation complex, thereby activating translation, possibly by positioning the 30S ribosomal subunit correctly for optimal translation initiation (Jagodnik et al., 2017). A ribosome stalled at ISL would lead to its following ribosome having increased initiation potential at the main ORF start codon, and perhaps even a queue behind that with elevated leader ORF starting potential (Ivanov et al., 2017). Disruption of the block when selenium (or perhaps toxins) became plentiful, could lead to a burst of downstream translation and potentially SELENOP synthesis. However, the ribosome profiling of living oysters shows only a 1.5-fold increase of main ORF translation under conditions of a 6-fold total selenium increase and western blot analysis on tissues with more elevated selenium does not show a dramatic increase (Supplementary Fig 4.5.3). Also, there is not a substantial increase of leader ORF protected fragments on selenium supplementation. In the case of oyster SelenoP, increased Se availability seems to function as an alternative switch at initiation, diminishing the ribosome accumulation observed when Se levels are increased possibly by disrupting ISL structure stability.

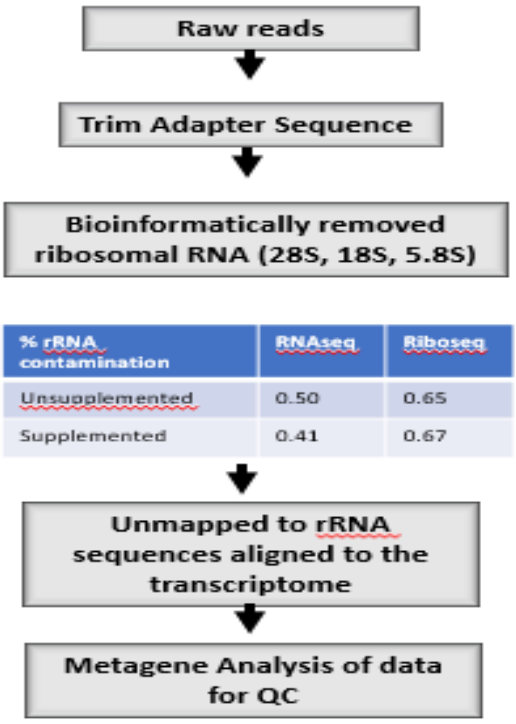
In conclusion, our results support a complex interplay of SelenoP mRNA elements with various factors to mediate SelenoP translation *in-vivo*. In addition to the slow decoding at UGA1 and processive multi-Sec incorporation of oyster SelenoP, we demonstrated a robust regulation at initiation possibly facilitated by RNA-elements and affected by Se-status. Although highly sensitive to Se-levels, the ISL-mediated effect seems to be independent of UGA 1 redefinition consistent with findings in vertebrates (Mariotti et al., 2017). It is also tempting to speculate that the lack of functional differentiation between oyster SECIS elements (as described in Chapter 3) improved its ability to support a highly processive Sec incorporation on its own SelenoP. However, caution is warranted in this interpretation as the oyster SECIS elements were examined in the context of zebrafish selenoP translation in a heterologous system and thus, is not an absolute reflection of *in vivo* mechanism.

Despite some interesting insights gained into the translation of oyster selenoP, further mechanistic questions are raised: what happens during this pause at initiation? How exactly does Se 'switch off' this initiation effect? Are there any other cis and trans-acting factors necessary to mediate processive Sec-incorporation *in-vivo*? Additional mechanistic insights at initiation-regulation could be gained by identification of factors that interact with ISL via cross-linking experiments. Further, it might also be informative to look at ribosome coverage of oysters cultivated in a Se-depleted environment, although such methods may pose a challenge like consequent depletion of other nutrients or ions necessary for survival (Vogel et al., 2018, Gore et al., 2010).

4.5 Supplementary Figures/Tables

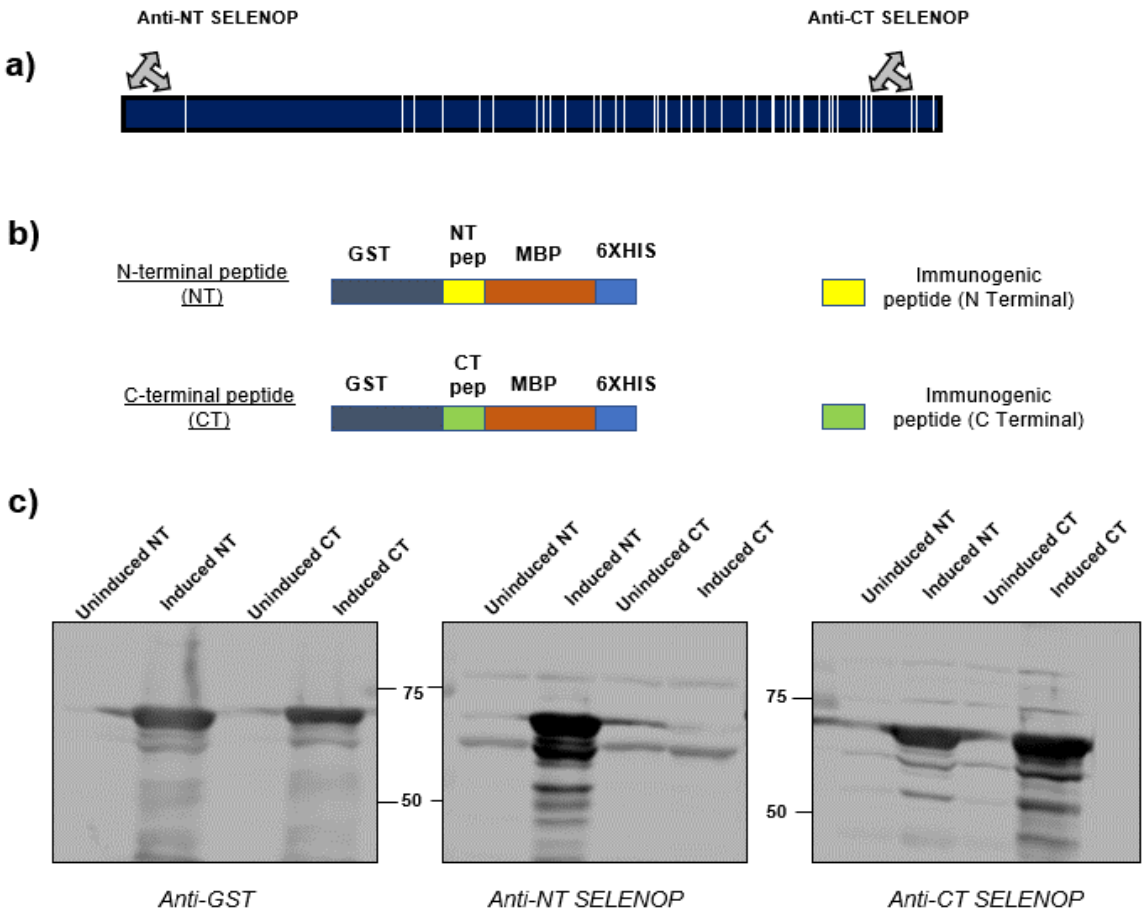


b)

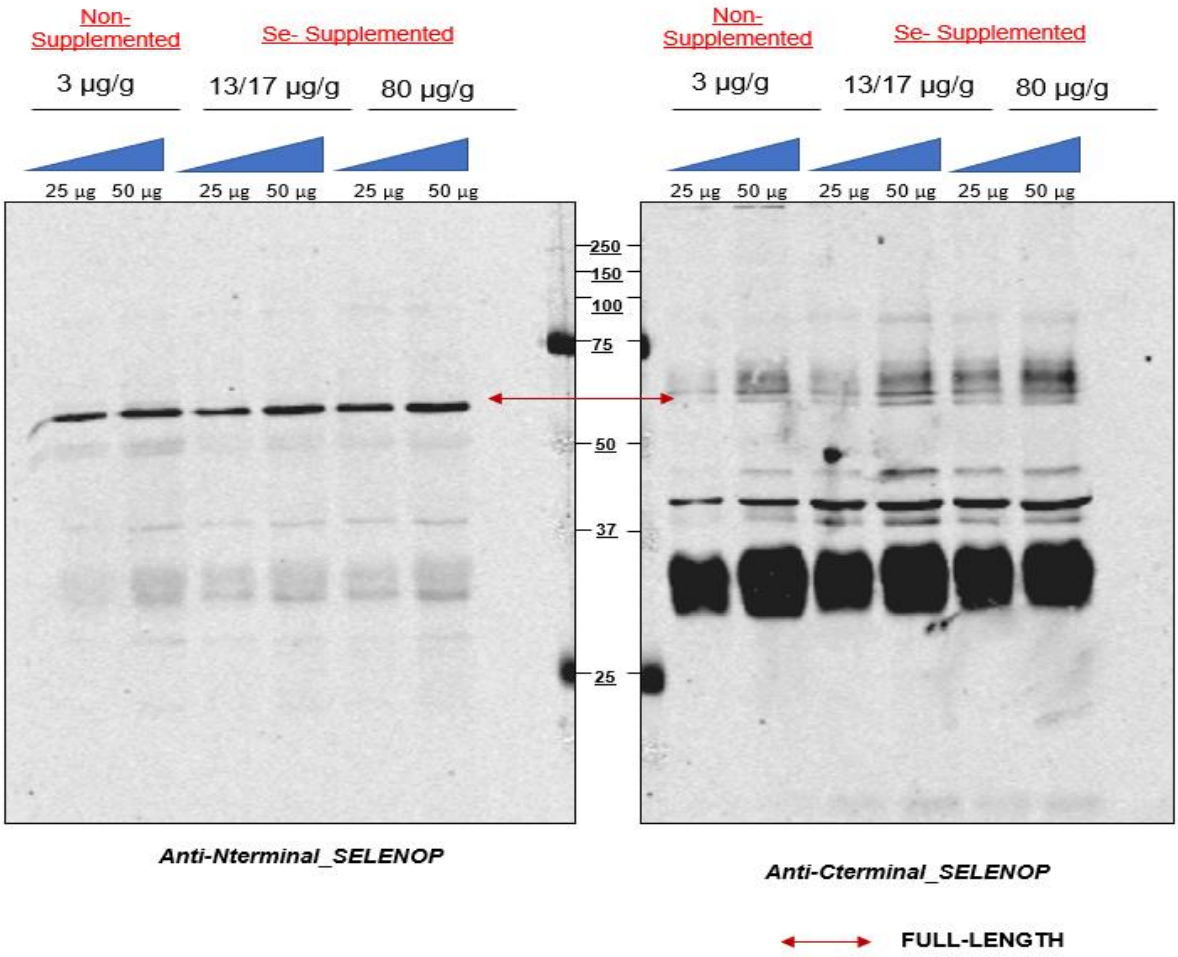


% rRNA contamination	RNAseq	Riboseq
Unsupplemented	0.50	0.65
Supplemented	0.41	0.67

Supplementary Figure 4.5.1: Schematics of ribosome profiling library generation and analysis. (a)Experimental work-flow of ribosome profiling and representative gels at each step. Green box denotes the product to be purified from each step of (i) Tissue lysis and RPF isolation (ii) 3'adapter ligation (iii) reverse transcription. The RT products were (v) circularized and (vi) PCR amplified at varying cycles. The top band was isolated for High-seq deep-sequencing. (b) Bioinformatic filters applied to raw reads. Ribosomal RNA (rRNA) contaminating reads were bioinformatically removed and percentage contamination are indicated for each sample of supplementation, RNA-seq and ribo-seq. rRNA sequences are listed in Suppl. Table 4.5.4.



Supplementary Figure 4.5.2: SELENOP Antibody Characterization.
(a) Protein map of SELENOP including Sec-UGAs (vertical white lines) and antibody recognition sites (grey) are indicated. N-terminal Ab (Anti-NT-SELENOP) is located before Sec - UGA1 and C-terminal Ab (Anti-CT-SELENOP) is located after Sec-UGA 43. (b) Antibody immunogenic peptide insertion: NT- N-terminal peptide (yellow) and CT C-terminal peptide (green) into a GST-MBP-6XHis (GST-MBP-NT-6XHis or GST-MBP-CT-6XHis) tagged vector for bacterial production. (c) Immunoblot of bacterial protein produced using antibody against the GST tag (Anti-GST), N-terminal SELENOP (Anti-NT-SelenoP) and C-terminal SELENOP (Anti-CT-SelenoP).



Supplementary Figure 4.5.3: Anti-SELENOP immunoblot of oyster tissues from supplementation trial with varying concentrations determined.

Oyster tissues with determined total Se levels were probed with antibodies against N and C terminal portions of SELENOP. Two total protein dilutions were loaded on the gel (25 and 50µg) of each sample:1 representative sample from the non-supplemented group and 2 representative samples from the Se-supplemented group with variable total Se levels determined.

General conclusions: Summary

The famous Jacob Monod phrase "Anything found to be true of *E. coli* must also be true of elephants" once reflected many biologists belief and understanding about the 'fixed' nature of the genetic code. We now know of many cases disproving this earlier assumption and therefore, such cases are in support of the genetic code's flexibility with implications to entire genomes or specific mRNAs. This work focused on two interesting cases of mRNA-specific genetic code expansion involving the UGA translation terminator codon in selenoproteins.

The first case described here is the Selenophosphate Synthetase 1 (SPS1-UGA) in the hymenopteran, *Apis mellifera*, void of the selenocysteine redefinition machinery (Chapter 2). Our attempts to identify amino-acid 'X' inserted in response to UGA involved endogenous SPS1 protein purification from a hymenopteran, *Apis mellifera*, source. We explored SPS1 protein dimerization properties to pull down endogenous SPS1 from honeybee lysates. Protein identification by mass spectrometry analysis revealed a peptide sequence specific to bee SPS1 located at the C-terminal portion of the protein indicative that the SPS1 mRNA is translated past the UGA codon. Despite unsuccessful efforts to identify the amino-acid inserted at this position, the work described in Chapter 2 laid a foundation of possible techniques to explore to be able to characterize the mechanism of translation of SPS1-UGA in hymenopterans. The self-dimerization property of SPS1 was an important technique that could be further optimized (e.g. with additional purification steps to avoid extensive proteolysis) to purify the endogenous protein. In addition, the bee primary cell line, with further developments, could prove to be an important tool for *sps1* gene introduction and subsequent exogenous protein expression. Supplementation of honey-bee-specific factors in cell-free translation systems, for instance, endogenous bee tRNAs or interacting proteins (Oudouhou et al., 2017), could provide some insights into elucidating the mechanism involved in this process. Nevertheless, there remains a huge gap in our knowledge of Sps1-UGA translation and thus, it is crucial to delineate interacting proteins or factors during Sps1-UGA translation in hymenopterans.

Selenoprotein P (SelenoP) presents an extreme case of deviation to the standard genetic decoding readout with the requirement to insert selenocysteines in response to multiple UGAs (10 in humans and rodents). In Chapter 3, the extended

search and phylogenetic construction of invertebrate SelenoP presented a remarkable diversity across metazoans of the region encoding the multi-Sec C-terminal domain. The origin of the Sec-rich tail pre-dates early metazoans (high Sec codon counts were identified in basal invertebrates like Cnidarians with 66 UGAs). The homology of SelenoP demonstrated across lineages show that this multi-Sec encoding tail contained repeated motifs similar to other occurrences within the same protein rather than to sequences encoded by SelenoP in other metazoan groups. This suggest that independent events of elongation occurred within taxonomic groups or lineages to extend the C-terminal tail. This invertebrate analysis of SelenoP also supported previous hypothesis (Lobanov et. al., 2007, Lobanov et al., 2008) associating aquatic organisms with increased selenium utilization as supported by a larger selenoproteome and more Sec residues in their SELENOP. It was shown that SELENOP evolution led to outstanding numbers of Sec-residues in metazoans with the record number of 132 UGAs found in the freshwater mussel, *Elliptio complanata*.

Fish and shellfish are known for their high content of selenium, with the mussel *Mytilus edulis* (blue oyster) scoring the highest selenium concentration in a recent survey across various marine organisms (Bryszewska and Mage, 2015). Yet, the role of selenium in the biology of these animals is virtually unexplored. In Chapter 4, we characterized the *Crassostrea gigas* selenoproteome in detail, and provided insights into Se effects on gene expression regulation *in-vivo*. We identified 32 selenoproteins in oyster, more than in mammals (25) (Mariotti et al., 2012). Our analysis highlighted interesting differences in the regulation of the oyster selenoproteome by dietary Se compared to mammals. These are not surprising, given their long phylogenetic distance and many differences in physiology. Upon supplementation, the remarkable increase in total selenium in their tissues (up to 50-fold) was accompanied by a more modest upregulation of selenoprotein levels. Thus, the levels of selenium in tissues cannot be accounted by just protein production, contrary to mammals where most selenium uptake is directed to selenocysteine-incorporation into selenoproteins (Schomburg and Schweizer, 2009, Kurokawa and Berry, 2013). This suggests the presence of additional molecular mechanisms for selenium accumulation, which may involve specialized protein machinery or non-protein components that bind selenium, e.g. low molecular weight selenocompounds or selenosugars. Altogether, our work paves the way for investigating selenium

biology in oysters and other molluscs. Considering the importance of these species in the food industry and their high selenium content, we expect that our research will prompt further studies to elucidate the function of selenoproteins and other aspects of selenium metabolism in these species.

The work in both Chapters 3 and 4 provided mechanistic insights into the translation of SelenoP, particularly in the *Crassostrea gigas* containing 46 UGAs. We identified a 16-codon leader ORF in *C.gigas* SelenoP whose UAA terminator is 22 nts 5' of the main ORF AUG start codon. The context of both the leader ORF and the main CDS have relatively poor Kozak sequence features. In addition, an RNA secondary structure termed Initiation Stem loop was identified 37-157 nt 3' of the main ORF start and 15 nt 5' of UGA 1. Its mammalian counterpart was shown to modulate translation initiation *in vitro* (Mariotti et. al., 2017). In oyster, the function of ISL appears to be selenium responsive. In fact, the oyster ribosome profiling shows an accumulation of abundant RPFs on the 5' side of the ISL, indicative of blockage of ribosome progression. The peak of RPFs is observed only in the non-supplemented samples, suggesting a regulatory mechanism sensing selenium levels. Whether the ISL functions directly as a potential selenium riboswitch is outside the scope of the present work. Does it merely act as a "gate" for ribosome progression to avoid wasteful downstream translation? Alternatively, is it part of a mechanism for programming ribosomes at initiation so that they later decode UGA as selenocysteine. While future studies involving selenium depletion are desirable, a paradoxical result was obtained with the heterologous *in vitro* translation experiments: synonymous codon substitution to disrupt the pairing involved in the ISL led to the absence of radiolabelled product (Chapter 3). However, current data do not allow us to distinguish whether this is due to abrogated initiation or inability to decode UGA as selenocysteine.

Our heterologous *in vitro* translation experiments revealed some degree of interchangeability in the Sec incorporation factors and SelenoP mRNA elements between invertebrates, fish and mammals. For instance, full-length translation of spider SelenoP with 9 Sec residues was achieved in rabbit reticulocyte lysates (RRL) with reconstituted rat CT-SECISBP2 indicating that all the factors necessary for its translation are present in this system. In contrast, while oyster SelenoP mRNA also

yielded Se labelled product, it was not full length and only a product corresponding to termination at UGAs 3 or 4. Replacement of the SECIS elements in zebrafish SelenoP mRNA with either both or single oyster counterparts and translation in reconstituted RRL showed that either oyster SECIS was equally able to support full-length translation (Chapter 3, Fig 3.3.15). With caution in extrapolating from this heterologous system, this may mean that invertebrate SelenoP SECIS 1 and 2 are more functionally interchangeable than their mammalian counterparts. Another example of likely divergence involves SECISBP2. Invertebrates have a single gene in this family and vertebrates have two, SECISBP2 and SECISBP2L. The former is a primary binder of SECIS elements and mouse conditional deletions gene revealed a significant reduction in mRNA-levels with retention of Sec specification (Seeher and Schweizer, 2014, Fradejas-Villar et al., 2017). The function of SECISBP2L is still unclear. Further research will be necessary to untangle the functions of the two paralogs and the differences with the oyster homolog seen in our heterologous experiments.

Irrespective of a role that ISL may potentially perform in programming ribosomes for UGA redefinition, what other features may be relevant to the programming? Since UGA (as well as UAG and UAA), is normally slow to decode by protein release factor(s), competition by cognate tRNA is likely facilitated. However, multiple features may also enhance these interactions. A known factor is the RNA structure at UGA1, SRE1, which is not required but relevant to efficiency (Mariotti et al. 2017). Another one is the overall selenoprotein mRNA structure enhancing SECIS complex proximity to UGA 1 (Turanov et al., 2009, Wu et al., 2016, Shetty and Copeland, 2018c) or interactions between the exon junction protein complex (EJC) and SECIS elements.

A likely constituent of the EJC, eIF4a3 is a selective negative regulator of selenoprotein synthesis and binds SECIS 2 (type I), but not SECIS 1 (type II) (Budiman et al., 2011). Notably, there is a conserved exon/intron junction 26 nts 3' of SelenoP UGA1 (of potential NMD relevance, this distance is smaller than 50 nt from UGA1; all the remaining UGAs are in the last exon). In one model, binding of eIF4a3 in the EJC to SelenoP SECIS 2, which is involved in redefinition of UGA1, may serve to facilitate localization of SECIS 2 for later replacement association with the oncoming ribosome. Such a potential mechanism cannot be obligatory since UGA1 can be recoded from mRNA generated from constructs lacking the intron, yet the high

level of termination observed with those constructs (Stoytcheva et al., 2006) suggests it may be relevant for improved efficiency.

Oyster ribosome profiling revealed inefficient (<5%) redefinition of UGA1 and approximating to 100% efficiency for the distal 3' UGAs (Chapter 4, Fig 4.3.10). In addition to the ribosome profiling results, the corresponding full-length translation of oyster SelenoP was demonstrated by metabolic Sec-incorporation into oyster SelenoP with ^{75}Se labelling of oyster larvae along with a parallel Western blot with anti-NT and anti-CT SELENOP. Given the larger numbers of distal UGAs in many cases and the proximity of these UGAs to each other (there are several occurrences of just one codon separating two UGA codons), we consider models involving the SECIS complex tracking with, or acting as a once off ribosome "switch" (review,(Atkins et al., 1999)) to be more appealing than separate contacts with a SECIS complex at each UGA. While there is good evidence from zebrafish to mammals for validity of the Berry model (UGA1 redefinition mainly enabled by SECIS 2, and redefinition of 3' UGAs by SECIS 1), we have no experimental data relevant to whether this model is applicable for invertebrates. However, our experiments showed that both SECIS elements of oyster SelenoP supported progressive Sec incorporation *in vitro* of a zebrafish construct, suggesting that such SECIS specialization is not present in oyster.

While there has been substantial debate about whether there are subsets of ribosomes free of mRNA that are specialized for translating specific mRNAs, decoding multiple UGAs in selenoprotein P as selenocysteine is the most striking example of mRNA-linked ribosome specialization. Ultimately, the resolution of highly processive translation of SelenoP will require substantial effort in delineating interactions of the essential cis- and trans- acting Sec-redefinition components with the ribosome combined with high-resolution structural studies.

To summarize, our work uncovered and explored the surprising flexibility of the genetic code across species. Such studies will provide important new insights into how the regulation of mRNA decoding fidelity relates to gene/mRNA compositions and how environmental cues shaped the evolution of some genomes. These examples further raise the possibility of engineering genomes of organisms capable of utilizing non-natural amino acids (Chin et al., 2003) to increase functional benefits of the resulting protein. Further, in the area of drug development, the increased reactivity of

selenocysteine is being explored to develop more potent and effective therapeutics (Li et al., 2017).

Publications

1. Wu, S. Marrioti M, Santesmasses D., Hill KE., **Baclaocos J,** Aparicio-Prat E., Li S., Mackrill J, Wu Y., Howard MT., Capecchi M., Guigo R, Burk RF and Atkins JF. Human selenoprotein P and S variant mRNAs with different numbers of SECIS elements and inferences from mutant mice of the roles of multiple SECIS elements. *Open biology* **6** (2016).

2. **Baclaocos, J.,** Santesmasses, D., Vetick M, Lynch S., Bierla K., Szpunar J., McAllen R., Mackrill J., Loughran G., Guigo R., Gladyshev VN, Mariotti M., and Atkins JF. 2019. SelenoP in Metazoa and Molluscs. *Journal of Molecular Biology*.

Bibliography

2010. Genome sequence of the pea aphid Acyrthosiphon pisum. *PLoS Biol,* 8**,** e1000313.

ABDELMOHSEN, K. & GOROSPE, M. 2012. RNA-binding protein nucleolin in disease. *RNA Biol,* 9**,** 799-808.

ALLMANG, C., CARBON, P. & KROL, A. 2002. The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *Rna,* 8**,** 1308-18.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology,* 215**,** 403-410.

ANAND, R., HOSKINS, A. A., STUBBE, J. & EALICK, S. E. 2004. Domain organization of Salmonella typhimurium formylglycinamide ribonucleotide amidotransferase revealed by X-ray crystallography. *Biochemistry,* 43**,** 10328-42.

ANDREEV, D. E., O'CONNOR, P. B., FAHEY, C., KENNY, E. M., TERENIN, I. M., DMITRIEV, S. E., CORMICAN, P., MORRIS, D. W., SHATSKY, I. N. & BARANOV, P. V. 2015. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife,* 4**,** e03971.

ANTONOV, I., COAKLEY, A., ATKINS, J. F., BARANOV, P. V. & BORODOVSKY, M. 2013. Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic acids research,* 41**,** 6514-6530.

ATKINS, J. F. & BARANOV, P. V. 2010. The distinction between recoding and codon reassignment. *Genetics,* 185**,** 1535-6.

ATKINS, J. F., BÖCK, A., MATSUFUJI, S. & GESTELAND, R. F. 1999. Dynamics of the genetic code. *COLD SPRING HARBOR MONOGRAPH SERIES,* 37**,** 637-674.

ATKINS, J. F. & GESTELAND, R. F. 2010. *Recoding : expansion of decoding rules enriches gene expression,* New York, Springer.

ATKINS, J. F., LOUGHRAN, G., BHATT, P. R., FIRTH, A. E. & BARANOV, P. V. 2016. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res,* 44**,** 7007-78.

BACLAOCOS, J., SANTESMASSES, D., MARIOTTI, M., BIERŁA, K., VETICK, M. B., LYNCH, S., MCALLEN, R., MACKRILL, J. J., LOUGHRAN, G., GUIGÓ, R., SZPUNAR, J., COPELAND, P. R., GLADYSHEV, V. N. & ATKINS, J. F. 2019. SelenoP in Metazoa and Molluscs. *Journal of Molecular Biology*.

BANERJEE, S., YANG, S. & FOSTER, C. B. 2012. A luciferase reporter assay to investigate the differential selenium-dependent stability of selenoprotein mRNAs. *The Journal of Nutritional Biochemistry,* 23**,** 1294-1301.

BANSAL, M. P., OBORN, C. J., DANIELSON, K. G. & MEDINA, D. 1989. Evidence for two selenium-binding proteins distinct from glutathione peroxidase in mouse liver. *Carcinogenesis,* 10**,** 541-546.

BARANOV, P. V., ATKINS, J. F. & YORDANOVA, M. M. 2015. Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. *Nat Rev Genet,* 16**,** 517-29.

BARANOV, P. V., GESTELAND, R. F. & ATKINS, J. F. 2002a. Recoding: translational bifurcations in gene expression. *Gene,* 286**,** 187-201.
BARANOV, P. V., GESTELAND, R. F. & ATKINS, J. F. 2002b. Release factor 2 frameshifting sites in different bacteria. *EMBO reports,* 3**,** 373.
BARANOV, P. V., HENDERSON, C. M., ANDERSON, C. B., GESTELAND, R. F., ATKINS, J. F. & HOWARD, M. T. 2005. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology,* 332**,** 498-510.
BARRELL, B. G., BANKIER, A. T. & DROUIN, J. 1979. A different genetic code in human mitochondria. *Nature,* 282**,** 189-94.
BARRY, J. K. & MILLER, W. A. 2002. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc Natl Acad Sci U S A,* 99**,** 11133-8.
BERRY, M. J., BANU, L., CHEN, Y. Y., MANDEL, S. J., KIEFFER, J. D., HARNEY, J. W. & LARSEN, P. R. 1991. Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature,* 353**,** 273-6.
BERRY, M. J., BANU, L., HARNEY, J. W. & LARSEN, P. R. 1993. Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *Embo j,* 12**,** 3315-22.
BIFANO, A. L., ATASSI, T., FERRARA, T. & DRISCOLL, D. M. 2013. Identification of nucleotides and amino acids that mediate the interaction between ribosomal protein L30 and the SECIS element. *BMC Mol Biol,* 14**,** 12.
BLANCHET, S., CORNU, D., ARGENTINI, M. & NAMY, O. 2014. *New insights into the incorporation of natural suppressor tRNAs at stop codons in Saccharomyces cerevisiae*.
BOCK, A. 2000. Biosynthesis of selenoproteins--an overview. *Biofactors,* 11**,** 77-8.
BÖCK, A., FORCHHAMMER, K., HEIDER, J., LEINFELDER, W., SAWERS, G., VEPREK, B. & ZINONI, F. 1991. Selenocysteine: the 21st amino acid. *Molecular Microbiology,* 5**,** 515-520.
BOSL, M. R., TAKAKU, K., OSHIMA, M., NISHIMURA, S. & TAKETO, M. M. 1997. Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (Trsp). *Proc Natl Acad Sci U S A,* 94**,** 5531-4.
BRAR, G. A., YASSOUR, M., FRIEDMAN, N., REGEV, A., INGOLIA, N. T. & WEISSMAN, J. S. 2012. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science,* 335**,** 552-7.
BRIERLEY, I., DIGARD, P. & INGLIS, S. C. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell,* 57**,** 537-47.
BROWN, A., SHAO, S., MURRAY, J., HEGDE, R. S. & RAMAKRISHNAN, V. 2015. Structural basis for stop codon recognition in eukaryotes. *Nature,* 524**,** 493-496.
BRULE, C. E. & GRAYHACK, E. J. 2017. Synonymous Codons: Choose Wisely for Expression. *Trends Genet,* 33**,** 283-297.
BRYSZEWSKA, M. A. & MAGE, A. 2015. Determination of selenium and its compounds in marine organisms. *J Trace Elem Med Biol,* 29**,** 91-8.
BUBENIK, J. L. & DRISCOLL, D. M. 2007. Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in

selenocysteine insertion sequence-binding protein 2. *J Biol Chem,* 282**,** 34653-62.
BUBENIK, J. L., MINIARD, A. C. & DRISCOLL, D. M. 2013. Alternative transcripts and 3'UTR elements govern the incorporation of selenocysteine into selenoprotein S. *PloS one,* 8**,** e62102-e62102.
BUDIMAN, M. E., BUBENIK, J. L. & DRISCOLL, D. M. 2011. Identification of a signature motif for the eIF4a3-SECIS interaction. *Nucleic Acids Res,* 39**,** 7730-9.
BUDIMAN, M. E., BUBENIK, J. L., MINIARD, A. C., MIDDLETON, L. M., GERBER, C. A., CASH, A. & DRISCOLL, D. M. 2009. Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. *Mol Cell,* 35**,** 479-89.
BURK, R. F. & HILL, K. E. 2009. Selenoprotein P-expression, functions, and roles in mammals. *Biochimica et biophysica acta,* 1790**,** 1441-1447.
BURK, R. F. & HILL, K. E. 2015. Regulation of Selenium Metabolism and Transport. *Annu Rev Nutr,* 35**,** 109-34.
BURK, R. F., HILL, K. E. & MOTLEY, A. K. 2001. Plasma selenium in specific and non-specific forms. *BioFactors,* 14**,** 107-114.
CABAN, K. & COPELAND, P. R. 2006. Size matters: a view of selenocysteine incorporation from the ribosome. *Cell Mol Life Sci,* 63**,** 73-81.
CABAN, K., KINZY, S. A. & COPELAND, P. R. 2007. The L7Ae RNA binding motif is a multifunctional domain required for the ribosome-dependent Sec incorporation activity of Sec insertion sequence binding protein 2. *Mol Cell Biol,* 27**,** 6350-60.
CAITO, S. W., MILATOVIC, D., HILL, K. E., ASCHNER, M., BURK, R. F. & VALENTINE, W. M. 2011. Progression of neurodegeneration and morphologic changes in the brains of juvenile mice with selenoprotein P deleted. *Brain Res,* 1398**,** 1-12.
CAPECCHI, M. R. 1967. Polypeptide chain termination in vitro: isolation of a release factor. *Proc Natl Acad Sci U S A,* 58**,** 1144-51.
CARLSON, B. A., LEE, B. J., TSUJI, P. A., TOBE, R., PARK, J. M., SCHWEIZER, U., GLADYSHEV, V. N. & HATFIELD, D. L. 2016. Selenocysteine tRNA[Ser]Sec: From Nonsense Suppressor tRNA to the Quintessential Constituent in Selenoprotein Biosynthesis. *In:* HATFIELD, D. L., SCHWEIZER, U., TSUJI, P. A. & GLADYSHEV, V. N. (eds.) *Selenium: Its Molecular Biology and Role in Human Health.* Cham: Springer International Publishing.
CARLSON, B. A., MOUSTAFA, M. E., SENGUPTA, A., SCHWEIZER, U., SHRIMALI, R., RAO, M., ZHONG, N., WANG, S., FEIGENBAUM, L., LEE, B. J., GLADYSHEV, V. N. & HATFIELD, D. L. 2007. Selective restoration of the selenoprotein population in a mouse hepatocyte selenoproteinless background with different mutant selenocysteine tRNAs lacking Um34. *J Biol Chem,* 282**,** 32591-602.
CARLSON, B. A., XU, X. M., GLADYSHEV, V. N. & HATFIELD, D. L. 2005. Selective rescue of selenoprotein expression in mice lacking a highly specialized methyl group in selenocysteine tRNA. *J Biol Chem,* 280**,** 5542-8.
CARLSON, B. A., XU, X. M., KRYUKOV, G. V., RAO, M., BERRY, M. J., GLADYSHEV, V. N. & HATFIELD, D. L. 2004. Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *Proc Natl Acad Sci U S A,* 101**,** 12848-53.

CARRILLO-TRIPP, J., DOLEZAL, A. G., GOBLIRSCH, M. J., MILLER, W. A., TOTH, A. L. & BONNING, B. C. 2016. In vivo and in vitro infection dynamics of honey bee viruses. *Sci Rep,* 6**,** 22265.

CASTELLANO, S., MOROZOVA, N., MOREY, M., BERRY, M. J., SERRAS, F., COROMINAS, M. & GUIGO, R. 2001. In silico identification of novel selenoproteins in the Drosophila melanogaster genome. *EMBO Rep,* 2**,** 697-702.

CHAMBERS, I., FRAMPTON, J., GOLDFARB, P., AFFARA, N., MCBAIN, W. & HARRISON, P. R. 1986. The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, TGA. *EMBO J,* 5**,** 1221-7.

CHAN, C. C., DOSTIE, J., DIEM, M. D., FENG, W., MANN, M., RAPPSILBER, J. & DREYFUSS, G. 2004. eIF4A3 is a novel component of the exon junction complex. *Rna,* 10**,** 200-9.

CHAN, M. M., CHOI, S. Y., CHAN, Q. W., LI, P., GUARNA, M. M. & FOSTER, L. J. 2010. Proteome profile and lentiviral transduction of cultured honey bee (Apis mellifera L.) cells. *Insect Mol Biol,* 19**,** 653-8.

CHAPPLE, C. E. & GUIGO, R. 2008. Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. *PLoS One,* 3**,** e2968.

CHAVATTE, L., BROWN, B. A. & DRISCOLL, D. M. 2005. Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat Struct Mol Biol,* 12**,** 408-16.

CHEN, G., CHANG, K. Y., CHOU, M. Y., BUSTAMANTE, C. & TINOCO, I., JR. 2009. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc Natl Acad Sci U S A,* 106**,** 12706-11.

CHIN, J. W., CROPP, T. A., ANDERSON, J. C., MUKHERJI, M., ZHANG, Z. & SCHULTZ, P. G. 2003. An expanded eukaryotic genetic code. *Science,* 301**,** 964-7.

CHITTUM, H. S., HILL, K. E., CARLSON, B. A., LEE, B. J., BURK, R. F. & HATFIELD, D. L. 1997. Replenishment of selenium deficient rats with selenium results in redistribution of the selenocysteine tRNA population in a tissue specific manner. *Biochim Biophys Acta,* 1359**,** 25-34.

CHOU, M. Y. & CHANG, K. Y. 2010. An intermolecular RNA triplex provides insight into structural determinants for the pseudoknot stimulator of -1 ribosomal frameshifting. *Nucleic Acids Res,* 38**,** 1676-85.

CLÉRY, A., BOURGUIGNON-IGEL, V., ALLMANG, C., KROL, A. & BRANLANT, C. 2007. An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic acids research,* 35**,** 1868-1884.

COMBS, G. F., JR. 2001. Selenium in global food systems. *Br J Nutr,* 85**,** 517-47.

COPELAND, P. R. & DRISCOLL, D. M. 1999. Purification, redox sensitivity, and RNA binding properties of SECIS-binding protein 2, a protein involved in selenoprotein biosynthesis. *J Biol Chem,* 274**,** 25447-54.

COPELAND, P. R., FLETCHER, J. E., CARLSON, B. A., HATFIELD, D. L. & DRISCOLL, D. M. 2000. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J,* 19**,** 306-14.

COPELAND, P. R., STEPANIK, V. A. & DRISCOLL, D. M. 2001. Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2. *Mol Cell Biol,* 21**,** 1491-8.

CRAIGEN, W. J. & CASKEY, C. T. 1986. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature,* 322**,** 273-5.

CRICK, F. H. 1968. The origin of the genetic code. *J Mol Biol,* 38**,** 367-79.

CRICK, F. H., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. 1961. General nature of the genetic code for proteins. *Nature,* 192**,** 1227-32.

CRIDGE, A. G., CROWE-MCAULIFFE, C., MATHEW, S. F. & TATE, W. P. 2018. Eukaryotic translational termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA channel. *Nucleic Acids Res,* 46**,** 1927-1944.

CUBADDA, F., AURELI, F., CIARDULLO, S., D'AMATO, M., RAGGI, A., ACHARYA, R., REDDY, R. A. V. & PRAKASH, N. T. 2010. Changes in Selenium Speciation Associated with Increasing Tissue Concentrations of Selenium in Wheat Grain. *Journal of Agricultural and Food Chemistry,* 58**,** 2295-2301.

DAEL, P. V., DAVIDSSON, L., MUÑOZ-BOX, R., FAY, L. B. & BARCLAY, D. 2007. Selenium absorption and retention from a selenite- or selenate-fortified milk-based formula in men measured by a stable-isotope technique. *British Journal of Nutrition,* 85**,** 157-163.

DEVER, T. E. & GREEN, R. 2012. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol,* 4**,** a013706.

DIAMOND, A. M., CHOI, I. S., CRAIN, P. F., HASHIZUME, T., POMERANTZ, S. C., CRUZ, R., STEER, C. J., HILL, K. E., BURK, R. F. & MCCLOSKEY, J. A. 1993. Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). *Journal of Biological Chemistry,* 268**,** 14215-14223.

DILWORTH, G. L. 1982. Properties of the selenium-containing moiety of nicotinic acid hydroxylase from Clostridium barkeri. *Arch Biochem Biophys,* 219**,** 30-8.

DOBOSZ-BARTOSZEK, M., PINKERTON, M. H., OTWINOWSKI, Z., CHAKRAVARTHY, S., SOLL, D., COPELAND, P. R. & SIMONOVIC, M. 2016. Crystal structures of the human elongation factor eEFSec suggest a non-canonical mechanism for selenocysteine incorporation. *Nat Commun,* 7**,** 12941.

DONOVAN, J., CABAN, K., RANAWEERA, R., GONZALEZ-FLORES, J. N. & COPELAND, P. R. 2008. A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment. *J Biol Chem,* 283**,** 35129-39.

DONOVAN, J. & COPELAND, P. R. 2009. Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins. *BMC Evol Biol,* 9**,** 229.

DONOVAN, J. & COPELAND, P. R. 2012. Selenocysteine insertion sequence binding protein 2L is implicated as a novel post-transcriptional regulator of selenoprotein expression. *PLoS One,* 7**,** e35581.

DREHER, T. W. & MILLER, W. A. 2006. Translational control in positive strand RNA plant viruses. *Virology,* 344**,** 185-97.

DUARTE, I., NABUURS, S. B., MAGNO, R. & HUYNEN, M. 2012. Evolution and diversification of the organellar release factor family. *Molecular biology and evolution,* 29**,** 3497-3512.

DUNN, J. G., FOO, C. K., BELLETIER, N. G., GAVIS, E. R. & WEISSMAN, J. S. 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *Elife,* 2**,** e01179.

EHRENREICH, A., FORCHHAMMER, K., TORMAY, P., VEPREK, B. & BOCK, A. 1992. Selenoprotein synthesis in E. coli. Purification and characterisation of the enzyme catalysing selenium activation. *Eur J Biochem,* 206**,** 767-73.

ENDOH, T. & SUGIMOTO, N. 2013. Unusual -1 ribosomal frameshift caused by stable RNA G-quadruplex in open reading frame. *Anal Chem,* 85**,** 11435-9.

ESAKI, N., NAKAMURA, T., TANAKA, H. & SODA, K. 1982. Selenocysteine lyase, a novel enzyme that specifically acts on selenocysteine. Mammalian distribution and purification and properties of pig liver enzyme. *J Biol Chem,* 257**,** 4386-91.

ESAKI, N., NAKAMURA, T., TANAKA, H., SUZUKI, T., MORINO, Y. & SODA, K. 1981. Enzymatic synthesis of selenocysteine in rat liver. *Biochemistry,* 20**,** 4492-6.

ESWARAPPA, S. M., POTDAR, A. A., KOCH, W. J., FAN, Y., VASU, K., LINDNER, D., WILLARD, B., GRAHAM, L. M., DICORLETO, P. E. & FOX, P. L. 2014. Programmed translational readthrough generates antiangiogenic VEGF-Ax. *Cell,* 157**,** 1605-18.

F. JR. COMBS, G. & LU, J. 2012. *Selenium as a cancer preventive agent*.

FAGEGALTIER, D., HUBERT, N., YAMADA, K., MIZUTANI, T., CARBON, P. & KROL, A. 2000a. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *Embo j,* 19**,** 4796-805.

FAGEGALTIER, D., LESCURE, A., WALCZAK, R., CARBON, P. & KROL, A. 2000b. Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res,* 28**,** 2679-89.

FINLEY, J. W. 2006. Bioavailability of Selenium from Foods. *Nutrition Reviews,* 64**,** 146-151.

FIRTH, A. E., WILLS, N. M., GESTELAND, R. F. & ATKINS, J. F. 2011. Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic Acids Res,* 39**,** 6679-91.

FIXSEN, S. M. & HOWARD, M. T. 2010. Processive selenocysteine incorporation during synthesis of eukaryotic selenoproteins. *J Mol Biol,* 399**,** 385-96.

FLETCHER, J. E., COPELAND, P. R. & DRISCOLL, D. M. 2000. Polysome distribution of phospholipid hydroperoxide glutathione peroxidase mRNA: evidence for a block in elongation at the UGA/selenocysteine codon. *Rna,* 6**,** 1573-84.

FRADEJAS-VILLAR, N., SEEHER, S., ANDERSON, C. B., DOENGI, M., CARLSON, B. A., HATFIELD, D. L., SCHWEIZER, U. & HOWARD, M. T. 2017. The RNA-binding protein Secisbp2 differentially modulates UGA codon reassignment and RNA decay. *Nucleic Acids Res,* 45**,** 4094-4107.

FROLOVA, L., LE GOFF, X., RASMUSSEN, H. H., CHEPEREGIN, S., DRUGEON, G., KRESS, M., ARMAN, I., HAENNI, A. L., CELIS, J. E., PHILIPPE, M. & ET AL. 1994. A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature,* 372**,** 701-3.

FÜßL, M., REINDERS, J., OEFNER, P., HEINZE, J. & SCHREMPF, A. 2014. *Selenophosphate-synthetase in the male accessory glands of an insect without selenoproteins*.
GANTHER, H. E. 1968. Selenotrisulfides. Formation by the reaction of thiols with selenious acid. *Biochemistry,* 7**,** 2898-905.
GANTHER, H. E. 1971. Reduction of the selenotrisulfide derivative of glutathione to a persulfide analog by gluthathione reductase. *Biochemistry,* 10**,** 4089-4098.
GARDIN, J., YEASMIN, R., YUROVSKY, A., CAI, Y., SKIENA, S. & FUTCHER, B. 2014. Measurement of average decoding rates of the 61 sense codons in vivo. *Elife,* 3.
GE, K., XUE, A., BAI, J. & WANG, S. 1983. Keshan disease-an endemic cardiomyopathy in China. *Virchows Arch A Pathol Anat Histopathol,* 401**,** 1-15.
GESTELAND, R. F., WEISS, R. B. & ATKINS, J. F. 1992. Recoding: reprogrammed genetic decoding. *Science,* 257**,** 1640-1.
GLADYSHEV, V. N., ARNER, E. S., BERRY, M. J., BRIGELIUS-FLOHE, R., BRUFORD, E. A., BURK, R. F., CARLSON, B. A., CASTELLANO, S., CHAVATTE, L., CONRAD, M., COPELAND, P. R., DIAMOND, A. M., DRISCOLL, D. M., FERREIRO, A., FLOHE, L., GREEN, F. R., GUIGO, R., HANDY, D. E., HATFIELD, D. L., HESKETH, J., HOFFMANN, P. R., HOLMGREN, A., HONDAL, R. J., HOWARD, M. T., HUANG, K., KIM, H. Y., KIM, I. Y., KOHRLE, J., KROL, A., KRYUKOV, G. V., LEE, B. J., LEE, B. C., LEI, X. G., LIU, Q., LESCURE, A., LOBANOV, A. V., LOSCALZO, J., MAIORINO, M., MARIOTTI, M., SANDEEP PRABHU, K., RAYMAN, M. P., ROZOVSKY, S., SALINAS, G., SCHMIDT, E. E., SCHOMBURG, L., SCHWEIZER, U., SIMONOVIC, M., SUNDE, R. A., TSUJI, P. A., TWEEDIE, S., URSINI, F., WHANGER, P. D. & ZHANG, Y. 2016. Selenoprotein Gene Nomenclature. *J Biol Chem,* 291**,** 24036-24040.
GLADYSHEV, V. N., KHANGULOV, S. V. & STADTMAN, T. C. 1994. Nicotinic acid hydroxylase from Clostridium barkeri: electron paramagnetic resonance studies show that selenium is coordinated with molybdenum in the catalytically active selenium-dependent enzyme. *Proc Natl Acad Sci U S A,* 91**,** 232-6.
GOBLER, C. J., LOBANOV, A. V., TANG, Y. Z., TURANOV, A. A., ZHANG, Y., DOBLIN, M., TAYLOR, G. T., SANUDO-WILHELMY, S. A., GRIGORIEV, I. V. & GLADYSHEV, V. N. 2013. The central role of selenium in the biochemistry and ecology of the harmful pelagophyte, Aureococcus anophagefferens. *Isme j,* 7**,** 1333-43.
GOBLIRSCH, M. J., SPIVAK, M. S. & KURTTI, T. J. 2013. A cell line resource derived from honey bee (Apis mellifera) embryonic tissues. *PLoS One,* 8**,** e69831.
GONZALEZ-FLORES, J. N., GUPTA, N., DEMONG, L. W. & COPELAND, P. R. 2012. The selenocysteine-specific elongation factor contains a novel and multi-functional domain. *J Biol Chem,* 287**,** 38936-45.
GOODY, T. A., MELCHER, S. E., NORMAN, D. G. & LILLEY, D. M. J. 2004. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA (New York, N.Y.),* 10**,** 254-264.
GORE, F., FAWELL, J. & BARTRAM, J. 2010. Too much or too little? A review of the conundrum of selenium. *Journal of water and health,* 8**,** 405-416.

GRUNDNER-CULEMANN, E., MARTIN, G. W., 3RD, HARNEY, J. W. & BERRY, M. J. 1999. Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *Rna,* 5**,** 625-35.

GU, Q. P., REAM, W. & WHANGER, P. D. 2002. Selenoprotein W gene regulation by selenium in L8 cells. *Biometals,* 15**,** 411-20.

GUIMARAES, M. J., PETERSON, D., VICARI, A., COCKS, B. G., COPELAND, N. G., GILBERT, D. J., JENKINS, N. A., FERRICK, D. A., KASTELEIN, R. A., BAZAN, J. F. & ZLOTNIK, A. 1996. Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism? *Proc Natl Acad Sci U S A,* 93**,** 15086-91.

GUPTA, N., DEMONG, L. W., BANDA, S. & COPELAND, P. R. 2013. Reconstitution of selenocysteine incorporation reveals intrinsic regulation by SECIS elements. *J Mol Biol,* 425**,** 2415-22.

GURVICH, O. L., NASVALL, S. J., BARANOV, P. V., BJORK, G. R. & ATKINS, J. F. 2011. Two groups of phenylalanine biosynthetic operon leader peptides genes: a high level of apparently incidental frameshifting in decoding Escherichia coli pheL. *Nucleic Acids Res,* 39**,** 3079-92.

HANDY, D. E. & LOSCALZO, J. 2012. Selenoproteins in Cardiovascular Redox Pathology. *In:* HATFIELD, D. L., BERRY, M. J. & GLADYSHEV, V. N. (eds.) *Selenium: Its Molecular Biology and Role in Human Health.* New York, NY: Springer New York.

HAO, B., ZHAO, G., KANG, P. T., SOARES, J. A., FERGUSON, T. K., GALLUCCI, J., KRZYCKI, J. A. & CHAN, M. K. 2004. Reactivity and chemical synthesis of L-pyrrolysine- the 22(nd) genetically encoded amino acid. *Chem Biol,* 11**,** 1317-24.

HARRELL, L., MELCHER, U. & ATKINS, J. F. 2002. Predominance of six different hexanucleotide recoding signals 3' of read-through stop codons. *Nucleic Acids Res,* 30**,** 2011-7.

HATFIELD, D., LEE, B. J., HAMPTON, L. & DIAMOND, A. M. 1991. Selenium induces changes in the selenocysteine tRNA[Ser]Sec population in mammalian cells. *Nucleic Acids Res,* 19**,** 939-43.

HATFIELD, D. L., TSUJI, P. A., CARLSON, B. A. & GLADYSHEV, V. N. 2014. Selenium and selenocysteine: roles in cancer, health, and development. *Trends Biochem Sci,* 39**,** 112-20.

HEDGE, L. H., KNOTT, N. A. & JOHNSTON, E. L. 2009. Dredging related metal bioaccumulation in oysters. *Marine Pollution Bulletin,* 58**,** 832-840.

HEROLD, J. & SIDDELL, S. G. 1993. An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res,* 21**,** 5838-42.

HERR, A. J., ATKINS, J. F. & GESTELAND, R. F. 2000a. Coupling of open reading frames by translational bypassing. *Annu Rev Biochem,* 69**,** 343-72.

HERR, A. J., GESTELAND, R. F. & ATKINS, J. F. 2000b. One protein from two open reading frames: mechanism of a 50 nt translational bypass. *Embo j,* 19**,** 2671-80.

HILL, K. E., LLOYD, R. S. & BURK, R. F. 1993. Conserved nucleotide sequences in the open reading frame and 3' untranslated region of selenoprotein P mRNA. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 537-541.

HILL, K. E., LLOYD, R. S., YANG, J. G., READ, R. & BURK, R. F. 1991. The cDNA for rat selenoprotein P contains 10 TGA codons in the open reading frame. *J Biol Chem,* 266**,** 10050-3.
HILL, K. E., ZHOU, J., MCMAHAN, W. J., MOTLEY, A. K., ATKINS, J. F., GESTELAND, R. F. & BURK, R. F. 2003. Deletion of selenoprotein P alters distribution of selenium in the mouse. *J Biol Chem,* 278**,** 13640-6.
HIMENO, S., CHITTUM, H. S. & BURK, R. F. 1996. Isoforms of selenoprotein P in rat plasma. Evidence for a full-length form and another form that terminates at the second UGA in the open reading frame. *J Biol Chem,* 271**,** 15769-75.
HINNEBUSCH, A. G., IVANOV, I. P. & SONENBERG, N. 2016. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science,* 352**,** 1413-1416.
HONDAL, R. J., MARINO, S. M. & GLADYSHEV, V. N. 2013. Selenocysteine in thiol/disulfide-like exchange reactions. *Antioxid Redox Signal,* 18**,** 1675-89.
HOROWITZ, S. & GOROVSKY, M. A. 1985. An unusual genetic code in nuclear genes of Tetrahymena. *Proc Natl Acad Sci U S A,* 82**,** 2452-5.
HOWARD, M. T., AGGARWAL, G., ANDERSON, C. B., KHATRI, S., FLANIGAN, K. M. & ATKINS, J. F. 2005. Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons. *Embo j,* 24**,** 1596-607.
HOWARD, M. T., CARLSON, B. A., ANDERSON, C. B. & HATFIELD, D. L. 2013. Translational redefinition of UGA codons is regulated by selenium availability. *J Biol Chem,* 288**,** 19401-13.
HOWARD, M. T., GESTELAND, R. F. & ATKINS, J. F. 2004. Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *Rna,* 10**,** 1653-61.
HOWARD, M. T., MOYLE, M. W., AGGARWAL, G., CARLSON, B. A. & ANDERSON, C. B. 2007. A recoding element that stimulates decoding of UGA codons by Sec tRNA[Ser]Sec. *Rna,* 13**,** 912-20.
HUANG, W. M., AO, S. Z., CASJENS, S., ORLANDI, R., ZEIKUS, R., WEISS, R., WINGE, D. & FANG, M. 1988. A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science,* 239**,** 1005-12.
HUBER, R. E. & CRIDDLE, R. S. 1967. Comparison of the chemical properties of selenocysteine and selenocystine with their sulfur analogs. *Arch Biochem Biophys,* 122**,** 164-73.
HUNTER, W. B. 2010. Medium for development of bee cell cultures (Apis mellifera: Hymenoptera: Apidae). *In Vitro Cell Dev Biol Anim,* 46**,** 83-6.
INGOLIA, N. T., BRAR, G. A., ROUSKIN, S., MCGEACHY, A. M. & WEISSMAN, J. S. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc,* 7**,** 1534-50.
INGOLIA, N. T., BRAR, G. A., STERN-GINOSSAR, N., HARRIS, M. S., TALHOUARNE, G. J., JACKSON, S. E., WILLS, M. R. & WEISSMAN, J. S. 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep,* 8**,** 1365-79.
INGOLIA, N. T., GHAEMMAGHAMI, S., NEWMAN, J. R. & WEISSMAN, J. S. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science,* 324**,** 218-23.

INGOLIA, N. T., LAREAU, L. F. & WEISSMAN, J. S. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell,* 147**,** 789-802.
ITOH, Y., CHIBA, S., SEKINE, S. & YOKOYAMA, S. 2009a. Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res,* 37**,** 6259-68.
ITOH, Y., SEKINE, S., MATSUMOTO, E., AKASAKA, R., TAKEMOTO, C., SHIROUZU, M. & YOKOYAMA, S. 2009b. Structure of selenophosphate synthetase essential for selenium incorporation into proteins and RNAs. *J Mol Biol,* 385**,** 1456-69.
IVANOV, I. P., GESTELAND, R. F. & ATKINS, J. F. 2000. Antizyme expression: a subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit. *Nucleic acids research,* 28**,** 3185-3196.
IVANOV, I. P., WEI, J., CASTER, S. Z., SMITH, K. M., MICHEL, A. M., ZHANG, Y., FIRTH, A. E., FREITAG, M., DUNLAP, J. C., BELL-PEDERSEN, D., ATKINS, J. F. & SACHS, M. S. 2017. Translation Initiation from Conserved Non-AUG Codons Provides Additional Layers of Regulation and Coding Capacity. *MBio,* 8.
JACKS, T. & VARMUS, H. E. 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science,* 230**,** 1237-42.
JAGODNIK, J., CHIARUTTINI, C. & GUILLIER, M. 2017. Stem-Loop Structures within mRNA Coding Sequences Activate Translation Initiation and Mediate Control by Small Regulatory RNAs. *Mol Cell,* 68**,** 158-170.e3.
JIANG, L., NI, J. & LIU, Q. 2012. Evolution of selenoproteins in the metazoan. *BMC Genomics,* 13**,** 446.
JU, H. & GHIL, S. 2015. Primary cell culture method for the honeybee Apis mellifera. *In Vitro Cell Dev Biol Anim,* 51**,** 890-3.
JUN, E., YE, J., HWANG, I., KIM, Y. & LEE, H. 2011. Selenium deficiency contributes to the chronic myocarditis in coxsackievirus-infected mice. *Acta virologica,* 55**,** 23-29.
JUNGREIS, I., LIN, M. F., SPOKONY, R., CHAN, C. S., NEGRE, N., VICTORSEN, A., WHITE, K. P. & KELLIS, M. 2011. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res,* 21**,** 2096-113.
KAUSHAL, N., GANDHI, U. H., NELSON, S. M., NARAYAN, V. & PRABHU, K. S. 2012. Selenium and Inflammation. *In:* HATFIELD, D. L., BERRY, M. J. & GLADYSHEV, V. N. (eds.) *Selenium: Its Molecular Biology and Role in Human Health.* New York, NY: Springer New York.
KAWAGUCHI, Y., HONDA, H., TANIGUCHI-MORIMURA, J. & IWASAKI, S. 1989. The codon CUG is read as serine in an asporogenic yeast Candida cylindracea. *Nature,* 341**,** 164-6.
KIM, H. K., LIU, F., FEI, J., BUSTAMANTE, C., GONZALEZ, R. L., JR. & TINOCO, I., JR. 2014. A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc Natl Acad Sci U S A,* 111**,** 5538-43.
KIM, H. Y. & GLADYSHEV, V. N. 2005. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol,* 3**,** e375.
KIM, I. Y., GUIMARAES, M. J., ZLOTNIK, A., BAZAN, J. F. & STADTMAN, T. C. 1997. Fetal mouse selenophosphate synthetase 2 (SPS2): characterization

of the cysteine mutant form overproduced in a baculovirus-insect cell system. *Proc Natl Acad Sci U S A,* 94**,** 418-21.
KINIRY, S. J., O'CONNOR, P. B. F., MICHEL, A. M. & BARANOV, P. V. 2019. Trips-Viz: a transcriptome browser for exploring Ribo-Seq data. *Nucleic Acids Res,* 47**,** D847-d852.
KIPP, A. P., STROHM, D., BRIGELIUS-FLOHE, R., SCHOMBURG, L., BECHTHOLD, A., LESCHIK-BONNET, E. & HESEKER, H. 2015. Revised reference values for selenium intake. *J Trace Elem Med Biol,* 32**,** 195-9.
KLAGGES, B. R., HEIMBECK, G., GODENSCHWEGE, T. A., HOFBAUER, A., PFLUGFELDER, G. O., REIFEGERSTE, R., REISCH, D., SCHAUPP, M., BUCHNER, S. & BUCHNER, E. 1996. Invertebrate synapsins: a single gene codes for several isoforms in Drosophila. *J Neurosci,* 16**,** 3154-65.
KLOBUTCHER, L. A. & FARABAUGH, P. J. 2002. Shifty ciliates: frequent programmed translational frameshifting in euplotids. *Cell,* 111**,** 763-6.
KNIGHT, R. D., FREELAND, S. J. & LANDWEBER, L. F. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet,* 2**,** 49-58.
KÖHRLE, J. 2000. The deiodinase family: selenoenzymes regulating thyroid hormone availability and action. *Cellular and Molecular Life Sciences CMLS,* 57**,** 1853-1863.
KOLLER, L. D. & EXON, J. H. 1986. The two faces of selenium-deficiency and toxicity--are similar in animals and man. *Can J Vet Res,* 50**,** 297-306.
KOROTKOV, K. V., NOVOSELOV, S. V., HATFIELD, D. L. & GLADYSHEV, V. N. 2002. Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element. *Mol Cell Biol,* 22**,** 1402-11.
KRYUKOV, G. V., CASTELLANO, S., NOVOSELOV, S. V., LOBANOV, A. V., ZEHTAB, O., GUIGO, R. & GLADYSHEV, V. N. 2003. Characterization of mammalian selenoproteomes. *Science,* 300**,** 1439-43.
KRYUKOV, G. V. & GLADYSHEV, V. N. 2004. The prokaryotic selenoproteome. *EMBO Rep,* 5**,** 538-43.
KUMAR, S., BJÖRNSTEDT, M. & HOLMGREN, A. 1992. Selenite is a substrate for calf thymus thioredoxin reductase and thioredoxin and elicits a large non-stoichiometric oxidation of NADPH in the presence of oxygen. *European Journal of Biochemistry,* 207**,** 435-439.
KUROKAWA, S. & BERRY, M. J. 2013. Selenium. Role of the essential metalloid in health. *Met Ions Life Sci,* 13**,** 499-534.
KUROKAWA, S., ERIKSSON, S., ROSE, K. L., WU, S., MOTLEY, A. K., HILL, S., WINFREY, V. P., MCDONALD, W. H., CAPECCHI, M. R., ATKINS, J. F., ARNER, E. S., HILL, K. E. & BURK, R. F. 2014. Sepp1(UF) forms are N-terminal selenoprotein P truncations that have peroxidase activity when coupled with thioredoxin reductase-1. *Free Radic Biol Med,* 69**,** 67-76.
LABUNSKYY, V. M., HATFIELD, D. L. & GLADYSHEV, V. N. 2014. Selenoproteins: molecular pathways and physiological roles. *Physiol Rev,* 94**,** 739-77.
LANG, B. F., JAKUBKOVA, M., HEGEDUSOVA, E., DAOUD, R., FORGET, L., BREJOVA, B., VINAR, T., KOSA, P., FRICOVA, D., NEBOHACOVA, M., GRIAC, P., TOMASKA, L., BURGER, G. & NOSEK, J. 2014. Massive programmed translational jumping in mitochondria. *Proc Natl Acad Sci U S A,* 111**,** 5926-31.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LARSEN, B., WILLS, N. M., GESTELAND, R. F. & ATKINS, J. F. 1994. rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J Bacteriol,* 176**,** 6842-51.

LATRECHE, L., JEAN-JEAN, O., DRISCOLL, D. M. & CHAVATTE, L. 2009. Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine. *Nucleic Acids Res,* 37**,** 5868-80.

LEE, B. J., RAJAGOPALAN, M., KIM, Y. S., YOU, K. H., JACOBSON, K. B. & HATFIELD, D. 1990. Selenocysteine tRNA[Ser]Sec gene is ubiquitous within the animal kingdom. *Mol Cell Biol,* 10**,** 1940-9.

LEE, K. H., SHIM, M. S., KIM, J. Y., JUNG, H. K., LEE, E., CARLSON, B. A., XU, X. M., PARK, J. M., HATFIELD, D. L., PARK, T. & LEE, B. J. 2011. Drosophila selenophosphate synthetase 1 regulates vitamin B6 metabolism: prediction and confirmation. *BMC Genomics,* 12**,** 426.

LEE, S., LIU, B., LEE, S., HUANG, S. X., SHEN, B. & QIAN, S. B. 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A,* 109**,** E2424-32.

LESCURE, A., GAUTHERET, D., CARBON, P. & KROL, A. 1999. Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. *J Biol Chem,* 274**,** 38147-54.

LESOON, A., MEHTA, A., SINGH, R., CHISOLM, G. M. & DRISCOLL, D. M. 1997. An RNA-binding protein recognizes a mammalian selenocysteine insertion sequence element required for cotranslational incorporation of selenocysteine. *Mol Cell Biol,* 17**,** 1977-85.

LI, C., KAPPOCK, T. J., STUBBE, J., WEAVER, T. M. & EALICK, S. E. 1999a. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the Escherichia coli purine biosynthetic pathway at 2.5 A resolution. *Structure,* 7**,** 1155-66.

LI, Q., IMATAKA, H., MORINO, S., ROGERS, G. W., JR., RICHTER-COOK, N. J., MERRICK, W. C. & SONENBERG, N. 1999b. Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol Cell Biol,* 19**,** 7336-46.

LI, X., NELSON, C. G., NAIR, R. R., HAZLEHURST, L., MORONI, T., MARTINEZ-ACEDO, P., NANNA, A. R., HYMEL, D., BURKE, T. R., JR. & RADER, C. 2017. Stable and Potent Selenomab-Drug Conjugates. *Cell Chem Biol,* 24**,** 433-442.e6.

LI, Y., TREFFERS, E. E., NAPTHINE, S., TAS, A., ZHU, L., SUN, Z., BELL, S., MARK, B. L., VAN VEELEN, P. A., VAN HEMERT, M. J., FIRTH, A. E., BRIERLEY, I., SNIJDER, E. J. & FANG, Y. 2014. Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc Natl Acad Sci U S A,* 111**,** E2172-81.

LIN, H. C., HO, S. C., CHEN, Y. Y., KHOO, K. H., HSU, P. H. & YEN, H. C. 2015. SELENOPROTEINS. CRL2 aids elimination of truncated selenoproteins produced by failed UGA/Sec decoding. *Science,* 349**,** 91-5.

LOBANOV, A. V., FOMENKO, D. E., ZHANG, Y., SENGUPTA, A., HATFIELD, D. L. & GLADYSHEV, V. N. 2007. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome biology,* 8**,** R198-R198.

LOBANOV, A. V., HATFIELD, D. L. & GLADYSHEV, V. N. 2008. Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biol,* 9**,** R62.

LOBANOV, A. V., HEAPHY, S. M., TURANOV, A. A., GERASHCHENKO, M. V., PUCCIARELLI, S., DEVARAJ, R. R., XIE, F., PETYUK, V. A., SMITH, R. D., KLOBUTCHER, L. A., ATKINS, J. F., MICELI, C., HATFIELD, D. L., BARANOV, P. V. & GLADYSHEV, V. N. 2017. Position-dependent termination and widespread obligatory frameshifting in Euplotes translation. *Nat Struct Mol Biol,* 24**,** 61-68.

LOBANOV, A. V., KRYUKOV, G. V., HATFIELD, D. L. & GLADYSHEV, V. N. 2006. Is there a twenty third amino acid in the genetic code? *Trends Genet,* 22**,** 357-60.

LONGSTAFF, D. G., BLIGHT, S. K., ZHANG, L., GREEN-CHURCH, K. B. & KRZYCKI, J. A. 2007. In vivo contextual requirements for UAG translation as pyrrolysine. *Mol Microbiol,* 63**,** 229-41.

LOUGHRAN, G., CHOU, M. Y., IVANOV, I. P., JUNGREIS, I., KELLIS, M., KIRAN, A. M., BARANOV, P. V. & ATKINS, J. F. 2014. Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res,* 42**,** 8928-38.

LOUGHRAN, G., HOWARD, M. T., FIRTH, A. E. & ATKINS, J. F. 2017. Avoidance of reporter assay distortions from fused dual reporters. *Rna,* 23**,** 1285-1289.

LOW, S. C., GRUNDNER-CULEMANN, E., HARNEY, J. W. & BERRY, M. J. 2000. SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy. *Embo j,* 19**,** 6882-90.

LU, J., ZHONG, L., LONN, M. E., BURK, R. F., HILL, K. E. & HOLMGREN, A. 2009. Penultimate selenocysteine residue replaced by cysteine in thioredoxin reductase from selenium-deficient rat liver. *Faseb j,* 23**,** 2394-402.

MA, S., HILL, K. E., CAPRIOLI, R. M. & BURK, R. F. 2002. Mass spectrometric characterization of full-length rat selenoprotein P and three isoforms shortened at the C terminus. Evidence that three UGA codons in the mRNA open reading frame have alternative functions of specifying selenocysteine insertion or translation termination. *J Biol Chem,* 277**,** 12749-54.

MARIOTTI, M. 2016. Selenocysteine Extinctions in Insects. *In:* RAMAN, C., GOLDSMITH, M. R. & AGUNBIADE, T. A. (eds.) *Short Views on Insect Genomics and Proteomics: Insect Proteomics, Vol.2.* Cham: Springer International Publishing.

MARIOTTI, M. & GUIGO, R. 2010. Selenoprofiles: profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *Bioinformatics,* 26**,** 2656-63.

MARIOTTI, M., LOBANOV, A. V., GUIGO, R. & GLADYSHEV, V. N. 2013. SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res,* 41**,** e149.

MARIOTTI, M., RIDGE, P. G., ZHANG, Y., LOBANOV, A. V., PRINGLE, T. H., GUIGO, R., HATFIELD, D. L. & GLADYSHEV, V. N. 2012. Composition

and Evolution of the Vertebrate and Mammalian Selenoproteomes. *PLOS ONE,* 7**,** e33066.
MARIOTTI, M., SANTESMASSES, D., CAPELLA-GUTIERREZ, S., MATEO, A., ARNAN, C., JOHNSON, R., D'ANIELLO, S., YIM, S. H., GLADYSHEV, V. N., SERRAS, F., COROMINAS, M., GABALDON, T. & GUIGO, R. 2015. Evolution of selenophosphate synthetases: emergence and relocation of function through independent duplications and recurrent subfunctionalization. *Genome Res,* 25**,** 1256-67.
MARIOTTI, M., SHETTY, S., BAIRD, L., WU, S., LOUGHRAN, G., COPELAND, P. R., ATKINS, J. F. & HOWARD, M. T. 2017. Multiple RNA structures affect translation initiation and UGA redefinition efficiency during synthesis of selenoprotein P. *Nucleic Acids Res,* 45**,** 13004-13015.
MATSUMOTO, E., SEKINE, S. I., AKASAKA, R., OTTA, Y., KATSURA, K., INOUE, M., KAMINISHI, T., TERADA, T., SHIROUZU, M. & YOKOYAMA, S. 2008. Structure of an N-terminally truncated selenophosphate synthetase from Aquifex aeolicus. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 64**,** 453-8.
MATSUMURA, S., IKAWA, Y. & INOUE, T. 2003. Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res,* 31**,** 5544-51.
MCCONNELL, K. P. & CHO, G. J. 1967. Active transport of L-selenomethionine in the intestine. *Am J Physiol,* 213**,** 150-6.
MEHTA, A., REBSCH, C. M., KINZY, S. A., FLETCHER, J. E. & COPELAND, P. R. 2004. Efficiency of mammalian selenocysteine incorporation. *J Biol Chem,* 279**,** 37852-9.
MERTZ, W. 1972. Human requirements: basic and optimal. *Ann N Y Acad Sci,* 199**,** 191-201.
MEYER, F., SCHMIDT, H. J., PLÜMPER, E., HASILIK, A., MERSMANN, G., MEYER, H. E., ENGSTRÖM, A. & HECKMANN, K. 1991. UGA is translated as cysteine in pheromone 3 of Euplotes octocarinatus. *Proceedings of the National Academy of Sciences,* 88**,** 3758-3761.
MINIARD, A. C., MIDDLETON, L. M., BUDIMAN, M. E., GERBER, C. A. & DRISCOLL, D. M. 2010. Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression. *Nucleic Acids Res,* 38**,** 4807-20.
MOGHADASZADEH, B., PETIT, N., JAILLARD, C., LINE, BROCKINGTON, M., QUIJANO-ROY, S., MERLINI, L., ROMERO, N., ESTOURNET, B., DESGUERRE, I., CHAIGNE, D., MUNTONI, F., TOPALOGLU, H. & GUICHENEY, P. 2001. *Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome*.
MORAR, M., ANAND, R., HOSKINS, A. A., STUBBE, J. & EALICK, S. E. 2006. Complexed structures of formylglycinamide ribonucleotide amidotransferase from Thermotoga maritima describe a novel ATP binding protein superfamily. *Biochemistry,* 45**,** 14880-95.
MORENO-REYES, R., MATHIEU, F., BOELAERT, M., BEGAUX, F., SUETENS, C., RIVERA, M. T., NEVE, J., PERLMUTTER, N. & VANDERPAS, J. 2003. Selenium and iodine supplementation of rural Tibetan children affected by Kashin-Beck osteoarthropathy. *Am J Clin Nutr,* 78**,** 137-44.
MOREY, M., COROMINAS, M. & SERRAS, F. 2003. DIAP1 suppresses ROS-induced apoptosis caused by impairment of the selD/sps1 homolog in Drosophila. *J Cell Sci,* 116**,** 4597-604.

MORIARTY, P. M., REDDY, C. C. & MAQUAT, L. E. 1998. Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. *Mol Cell Biol,* 18**,** 2932-9.

MOXON, A. L. & RHIAN, M. 1943. Selenium poisoning. *Physiological Reviews,* 23**,** 305-337.

MUHLHAUSEN, S., FINDEISEN, P., PLESSMANN, U., URLAUB, H. & KOLLMAR, M. 2016. A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Res,* 26**,** 945-55.

MUKAI, T., ENGLERT, M., TRIPP, H. J., MILLER, C., IVANOVA, N. N., RUBIN, E. M., KYRPIDES, N. C. & SOLL, D. 2016. Facile Recoding of Selenocysteine in Nature. *Angew Chem Int Ed Engl,* 55**,** 5337-41.

MULLAN, J. P. A., VELAYUDHAN, V., O'CONNOR, P. B. F., DONOHUE, C. A. & BARANOV, P. V. 2016. RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data AU - Michel, Audrey M. *RNA Biology,* 13**,** 316-319.

NAMY, O., DUCHATEAU-NGUYEN, G., HATIN, I., HERMANN-LE DENMAT, S., TERMIER, M. & ROUSSET, J. P. 2003. Identification of stop codon readthrough genes in Saccharomyces cerevisiae. *Nucleic Acids Res,* 31**,** 2289-96.

NAMY, O., HATIN, I. & ROUSSET, J. P. 2001. Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO reports,* 2**,** 787-793.

NAMY, O. & ROUSSET, J.-P. 2010. Specification of Standard Amino Acids by Stop Codons. *In:* ATKINS, J. F. & GESTELAND, R. F. (eds.) *Recoding: Expansion of Decoding Rules Enriches Gene Expression.* New York, NY: Springer New York.

NAMY, O., ZHOU, Y., GUNDLLAPALLI, S., POLYCARPO, C. R., DENISE, A., ROUSSET, J. P., SOLL, D. & AMBROGELLY, A. 2007. Adding pyrrolysine to the Escherichia coli genetic code. *FEBS Lett,* 581**,** 5282-8.

NASIM, M. T., JAENECKE, S., BELDUZ, A., KOLLMUS, H., FLOHE, L. & MCCARTHY, J. E. 2000. Eukaryotic selenocysteine incorporation follows a nonprocessive mechanism that competes with translational termination. *J Biol Chem,* 275**,** 14846-52.

NAVARRO-ALARCON, M. & CABRERA-VIQUE, C. 2008. Selenium in food and the human body: A review. *Science of The Total Environment,* 400**,** 115-141.

NEUHIERL, B., THANBICHLER, M., LOTTSPEICH, F. & BOCK, A. 1999. A family of S-methylmethionine-dependent thiol/selenol methyltransferases. Role in selenium tolerance and evolutionary relation. *J Biol Chem,* 274**,** 5407-14.

NIRENBERG, M., CASKEY, T., MARSHALL, R., BRIMACOMBE, R., KELLOGG, D., DOCTOR, B., HATFIELD, D., LEVIN, J., ROTTMAN, F., PESTKA, S., WILCOX, M. & ANDERSON, F. 1966. The RNA code and protein synthesis. *Cold Spring Harb Symp Quant Biol,* 31**,** 11-24.

NIRENBERG, M. W., MATTHAEI, J. H., JONES, O. W., MARTIN, R. G. & BARONDES, S. H. 1963. Approximation of genetic code via cell-free protein synthesis directed by template RNA. *Fed Proc,* 22**,** 55-61.

OGRA, Y., ISHIWATA, K., RUIZ ENCINAR, J., ŁOBIŃSKI, R. & SUZUKI, K. T. 2004. Speciation of selenium in selenium-enriched shiitake mushroom, Lentinula edodes. *Analytical and Bioanalytical Chemistry,* 379**,** 861-866.

OKUNO, T., MOTOBAYASHI, S. J., UENO, H. & NAKAMURO, K. 2005. Purification and characterization of mouse hepatic enzyme that converts selenomethionine to methylselenol by its α,γ-elimination. *Biological Trace Element Research,* 106**,** 77-93.

OLSON, G. E., WINFREY, V. P., NAGDAS, S. K., HILL, K. E. & BURK, R. F. 2005. Selenoprotein P is required for mouse sperm development. *Biol Reprod,* 73**,** 201-11.

OLSON, O. E., NOVACEK, E. J., WHITEHEAD, E. I. & PALMER, I. S. 1970. Investigations on selenium in wheat. *Phytochemistry,* 9**,** 1181-1188.

OLSTHOORN, R. C., LAURS, M., SOHET, F., HILBERS, C. W., HEUS, H. A. & PLEIJ, C. W. 2004. Novel application of sRNA: stimulation of ribosomal frameshifting. *Rna,* 10**,** 1702-3.

OUDOUHOU, F., CASU, B., DOPGWA PUEMI, A. S., SYGUSCH, J. & BARON, C. 2017. Analysis of Novel Interactions between Components of the Selenocysteine Biosynthesis Pathway, SEPHS1, SEPHS2, SEPSECS, and SECp43. *Biochemistry,* 56**,** 2261-2270.

PAPP, L. V., LU, J., STRIEBEL, F., KENNEDY, D., HOLMGREN, A. & KHANNA, K. K. 2006. The redox state of SECIS binding protein 2 controls its localization and selenocysteine incorporation function. *Mol Cell Biol,* 26**,** 4895-910.

PEDERSEN, J. S., BEJERANO, G., SIEPEL, A., ROSENBLOOM, K., LINDBLAD-TOH, K., LANDER, E. S., KENT, J., MILLER, W. & HAUSSLER, D. 2006. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol,* 2**,** e33.

PERSSON, B. C., BOCK, A., JACKLE, H. & VORBRUGGEN, G. 1997. SelD homolog from Drosophila lacking selenide-dependent monoselenophosphate synthetase activity. *J Mol Biol,* 274**,** 174-80.

PESOLE, G., LOTTI, M., ALBERGHINA, L. & SACCONE, C. 1995. Evolutionary origin of nonuniversal CUGSer codon in some Candida species as inferred from a molecular phylogeny. *Genetics,* 141**,** 903-907.

PLANT, E. P., PEREZ-ALVARADO, G. C., JACOBS, J. L., MUKHOPADHYAY, B., HENNIG, M. & DINMAN, J. D. 2005. A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol,* 3**,** e172.

PRAT, L., HEINEMANN, I. U., AERNI, H. R., RINEHART, J., O'DONOGHUE, P. & SOLL, D. 2012. Carbon source-dependent expansion of the genetic code in bacteria. *Proc Natl Acad Sci U S A,* 109**,** 21070-5.

PREER, J. R., JR., PREER, L. B., RUDMAN, B. M. & BARNETT, A. J. 1985. Deviation from the universal code shown by the gene for surface protein 51A in Paramecium. *Nature,* 314**,** 188-90.

PRERE, M. F., CANAL, I., WILLS, N. M., ATKINS, J. F. & FAYET, O. 2011. The interplay of mRNA stimulatory signals required for AUU-mediated initiation and programmed -1 ribosomal frameshifting in decoding of transposable element IS911. *J Bacteriol,* 193**,** 2735-44.

RAINBOW, P. S. & SMITH, B. D. 2010. Trophic transfer of trace metals: Subcellular compartmentalisation in bivalve prey and comparative assimilation efficiencies of two invertebrate predators. *Journal of Experimental Marine Biology and Ecology,* 390**,** 143-148.

RAMOS, S., FAQUIN, V., GUILHERME, L., CASTRO, E., ÁVILA, F., CARVALHO, G., BASTOS, C. & OLIVEIRA, C. 2010. Selenium

biofortification and antioxidant activity in lettuce plants fed with selenate and selenite. *Plant soil environ,* 56**,** 584-588.
RAYMAN, M. P. 2012. Selenium and human health. *Lancet,* 379**,** 1256-68.
RAZGA, F., KOCA, J., SPONER, J. & LEONTIS, N. B. 2005. Hinge-like motions in RNA kink-turns: the role of the second a-minor motif and nominally unpaired bases. *Biophys J,* 88**,** 3466-85.
READ, R., BELLEW, T., YANG, J. G., HILL, K. E., PALMER, I. S. & BURK, R. F. 1990. Selenium and amino acid composition of selenoprotein P, the major selenoprotein in rat serum. *J Biol Chem,* 265**,** 17899-905.
REEVES, M. A. & HOFFMANN, P. R. 2009. The human selenoproteome: recent insights into functions and regulation. *Cell Mol Life Sci,* 66**,** 2457-78.
RILEY, R., HARIDAS, S., WOLFE, K. H., LOPES, M. R., HITTINGER, C. T., GOKER, M., SALAMOV, A. A., WISECAVER, J. H., LONG, T. M., CALVEY, C. H., AERTS, A. L., BARRY, K. W., CHOI, C., CLUM, A., COUGHLAN, A. Y., DESHPANDE, S., DOUGLASS, A. P., HANSON, S. J., KLENK, H. P., LABUTTI, K. M., LAPIDUS, A., LINDQUIST, E. A., LIPZEN, A. M., MEIER-KOLTHOFF, J. P., OHM, R. A., OTILLAR, R. P., PANGILINAN, J. L., PENG, Y., ROKAS, A., ROSA, C. A., SCHEUNER, C., SIBIRNY, A. A., SLOT, J. C., STIELOW, J. B., SUN, H., KURTZMAN, C. P., BLACKWELL, M., GRIGORIEV, I. V. & JEFFRIES, T. W. 2016. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A,* 113**,** 9882-7.
RIVIERE, G., KLOPP, C., IBOUNIYAMINE, N., HUVET, A., BOUDRY, P. & FAVREL, P. 2015. GigaTON: an extensive publicly searchable database providing a new reference transcriptome in the pacific oyster Crassostrea gigas. *BMC Bioinformatics,* 16**,** 401.
RIYASATY, S. & ATKINS, J. F. 1968. External suppression of a frameshift mutant in salmonella. *J Mol Biol,* 34**,** 541-57.
ROBINSON, D. N. & COOLEY, L. 1997. Examination of the function of two kelch proteins generated by stop codon suppression. *Development,* 124**,** 1405-17.
SABBAGH, M. & VAN HOEWYK, D. 2012. Malformed Selenoproteins Are Removed by the Ubiquitin–Proteasome Pathway in Stanleya pinnata. *Plant and Cell Physiology,* 53**,** 555-564.
SAMSON, M. L., LISBIN, M. J. & WHITE, K. 1995. Two distinct temperature-sensitive alleles at the elav locus of Drosophila are suppressed nonsense mutations of the same tryptophan codon. *Genetics,* 141**,** 1101-11.
SARANGI, G. K., ROMAGNE, F. & CASTELLANO, S. 2018. Distinct Patterns of Selection in Selenium-Dependent Genes between Land and Aquatic Vertebrates. *Mol Biol Evol,* 35**,** 1744-1756.
SCHILD, F., KIEFFER-JAQUINOD, S., PALENCIA, A., COBESSI, D., SARRET, G., ZUBIETA, C., JOURDAIN, A., DUMAS, R., FORGE, V., TESTEMALE, D., BOURGUIGNON, J. & HUGOUVIEUX, V. 2014. Biochemical and biophysical characterization of the selenium-binding and reducing site in Arabidopsis thaliana homologue to mammals selenium-binding protein 1. *The Journal of biological chemistry,* 289**,** 31765-31776.
SCHOMBURG, L. & SCHWEIZER, U. 2009. Hierarchical regulation of selenoprotein expression and sex-specific effects of selenium. *Biochim Biophys Acta,* 1790**,** 1453-62.

SCHOMBURG, L., SCHWEIZER, U., HOLTMANN, B., FLOHE, L., SENDTNER, M. & KOHRLE, J. 2003. Gene disruption discloses role of selenoprotein P in selenium delivery to target tissues. *Biochem J,* 370**,** 397-402.
SCHUEREN, F., LINGNER, T., GEORGE, R., HOFHUIS, J., DICKEL, C., GARTNER, J. & THOMS, S. 2014. Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife,* 3**,** e03640.
SCHULTE, C., LEBOULLE, G., OTTE, M., GRUNEWALD, B., GEHNE, N. & BEYE, M. 2013. Honey bee promoter sequences for targeted gene expression. *Insect Mol Biol,* 22**,** 399-410.
SCOLNICK, E., TOMPKINS, R., CASKEY, T. & NIRENBERG, M. 1968. Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci U S A,* 61**,** 768-74.
SEEHER, S., ATASSI, T., MAHDI, Y., CARLSON, B. A., BRAUN, D., WIRTH, E. K., KLEIN, M. O., REIX, N., MINIARD, A. C., SCHOMBURG, L., HATFIELD, D. L., DRISCOLL, D. M. & SCHWEIZER, U. 2014. Secisbp2 is essential for embryonic development and enhances selenoprotein expression. *Antioxid Redox Signal,* 21**,** 835-49.
SEEHER, S. & SCHWEIZER, U. 2014. Targeted deletion of Secisbp2 reduces, but does not abrogate, selenoprotein expression and leads to striatal interneuron loss. *Free Radic Biol Med,* 75 Suppl 1**,** S9.
SELF, W. T. & STADTMAN, T. C. 2000. Selenium-dependent metabolism of purines: A selenium-dependent purine hydroxylase and xanthine dehydrogenase were purified from Clostridium purinolyticum and characterized. *Proc Natl Acad Sci U S A,* 97**,** 7208-13.
SHALGI, R., HURT, J. A., KRYKBAEVA, I., TAIPALE, M., LINDQUIST, S. & BURGE, C. B. 2013. Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell,* 49**,** 439-52.
SHETTY, S. & COPELAND, P. R. 2018a. Molecular mechanism of selenoprotein P synthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1862**,** 2506-2510.
SHETTY, S. & COPELAND, P. R. 2018b. Molecular mechanism of selenoprotein P synthesis. *Biochim Biophys Acta*.
SHETTY, S. P. & COPELAND, P. R. 2018c. The Selenium Transport Protein, Selenoprotein P, Requires Coding Sequence Determinants to Promote Efficient Selenocysteine Incorporation. *J Mol Biol*.
SHETTY, S. P., SHAH, R. & COPELAND, P. R. 2014. Regulation of selenocysteine incorporation into the selenium transport protein, selenoprotein P. *J Biol Chem,* 289**,** 25317-26.
SHETTY, S. P., STURTS, R. J., VETICK, M. B. & COPELAND, P. R. 2018. Processive incorporation of multiple selenocysteine residues is driven by a novel feature of the selenocysteine insertion sequence. *J Biol Chem*.
SHIBATA, N. & TORAYA, T. 2015. Molecular architectures and functions of radical enzymes and their (re)activating proteins. *J Biochem,* 158**,** 271-92.
SKUZESKI, J. M., NICHOLS, L. M., GESTELAND, R. F. & ATKINS, J. F. 1991. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *J Mol Biol,* 218**,** 365-73.
SMALL-HOWARD, A., MOROZOVA, N., STOYTCHEVA, Z., FORRY, E. P., MANSELL, J. B., HARNEY, J. W., CARLSON, B. A., XU, X. M., HATFIELD, D. L. & BERRY, M. J. 2006. Supramolecular complexes mediate selenocysteine incorporation in vivo. *Mol Cell Biol,* 26**,** 2337-46.

SNIDER, G. W., RUGGLES, E., KHAN, N. & HONDAL, R. J. 2013. Selenocysteine confers resistance to inactivation by oxidation in thioredoxin reductase: comparison of selenium and sulfur enzymes. *Biochemistry,* 52**,** 5472-81.

SONENBERG, N. & HINNEBUSCH, A. G. 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell,* 136**,** 731-45.

SONGE-MOLLER, L., VAN DEN BORN, E., LEIHNE, V., VAGBO, C. B., KRISTOFFERSEN, T., KROKAN, H. E., KIRPEKAR, F., FALNES, P. O. & KLUNGLAND, A. 2010. Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. *Mol Cell Biol,* 30**,** 1814-27.

SORS, T. G., ELLIS, D. R. & SALT, D. E. 2005. Selenium uptake, translocation, assimilation and metabolic fate in plants. *Photosynthesis Research,* 86**,** 373-389.

SPILLER, H. A. 2018. Rethinking mercury: the role of selenium in the pathophysiology of mercury toxicity. *Clin Toxicol (Phila),* 56**,** 313-326.

SQUIRES, J. E., STOYTCHEV, I., FORRY, E. P. & BERRY, M. J. 2007. SBP2 binding affinity is a major determinant in differential selenoprotein mRNA translation and sensitivity to nonsense-mediated decay. *Mol Cell Biol,* 27**,** 7848-55.

SRINIVASAN, G., JAMES, C. M. & KRZYCKI, J. A. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science,* 296**,** 1459-62.

STEINBRENNER, H., SPECKMANN, B. & KLOTZ, L.-O. 2016. Selenoproteins: Antioxidant selenoenzymes and beyond. *Archives of Biochemistry and Biophysics,* 595**,** 113-119.

STEINMANN, D., NAUSER, T. & KOPPENOL, W. H. 2010. Selenium and sulfur in exchange reactions: a comparative study. *J Org Chem,* 75**,** 6696-9.

STEITZ, J. A. 1969. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature,* 224**,** 957-64.

STENEBERG, P. & SAMAKOVLIS, C. 2001. A novel stop codon readthrough mechanism produces functional Headcase protein in Drosophila trachea. *EMBO Rep,* 2**,** 593-7.

STERN-GINOSSAR, N., WEISBURD, B., MICHALSKI, A., LE, V. T., HEIN, M. Y., HUANG, S. X., MA, M., SHEN, B., QIAN, S. B., HENGEL, H., MANN, M., INGOLIA, N. T. & WEISSMAN, J. S. 2012. Decoding human cytomegalovirus. *Science,* 338**,** 1088-93.

STIEBLER, A. C., FREITAG, J., SCHINK, K. O., STEHLIK, T., TILLMANN, B. A., AST, J. & BOLKER, M. 2014. Ribosomal readthrough at a short UGA stop codon context triggers dual localization of metabolic enzymes in Fungi and animals. *PLoS Genet,* 10**,** e1004685.

STOYTCHEVA, Z., TUJEBAJEVA, R. M., HARNEY, J. W. & BERRY, M. J. 2006. Efficient incorporation of multiple selenocysteines involves an inefficient decoding step serving as a potential translational checkpoint and ribosome bottleneck. *Mol Cell Biol,* 26**,** 9177-84.

STURCHLER, C., WESTHOF, E., CARBON, P. & KROL, A. 1993. Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic Acids Res,* 21**,** 1073-9.

SUN, X., ZHONG, Y., HUANG, Z. & YANG, Y. 2014. Selenium accumulation in unicellular green alga Chlorella vulgaris and its effects on antioxidant enzymes and content of photosynthetic pigments. *PLoS One,* 9**,** e112270.
SUNDE, R. A. 2010. mRNA transcripts as molecular biomarkers in medicine and nutrition. *The Journal of nutritional biochemistry,* 21**,** 665-670.
SUNDE, R. A. & RAINES, A. M. 2011. Selenium regulation of the selenoprotein and nonselenoprotein transcriptomes in rodents. *Adv Nutr,* 2**,** 138-50.
SUZUKI, K. T., KURASAKI, K. & SUZUKI, N. 2007. Selenocysteine beta-lyase and methylselenol demethylase in the metabolism of Se-methylated selenocompounds into selenide. *Biochim Biophys Acta,* 1770**,** 1053-61.
TAJIMA, Y., IWAKAWA, H. O., KAIDO, M., MISE, K. & OKUNO, T. 2011. A long-distance RNA-RNA interaction plays an important role in programmed -1 ribosomal frameshifting in the translation of p88 replicase protein of Red clover necrotic mosaic virus. *Virology,* 417**,** 169-78.
TAKEUCHI, A., SCHMITT, D., CHAPPLE, C., BABAYLOVA, E., KARPOVA, G., GUIGO, R., KROL, A. & ALLMANG, C. 2009. A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins. *Nucleic Acids Res,* 37**,** 2126-41.
TAMURA, T., YAMAMOTO, S., TAKAHATA, M., SAKAGUCHI, H., TANAKA, H., STADTMAN, T. C. & INAGAKI, K. 2004. Selenophosphate synthetase genes from lung adenocarcinoma cells: Sps1 for recycling L-selenocysteine and Sps2 for selenite assimilation. *Proc Natl Acad Sci U S A,* 101**,** 16162-7.
TOUAT-HAMICI, Z., LEGRAIN, Y., BULTEAU, A. L. & CHAVATTE, L. 2014. Selective up-regulation of human selenoproteins in response to oxidative stress. *J Biol Chem,* 289**,** 14750-61.
TSUJI, P. A., CARLSON, B. A., ANDERSON, C. B., SEIFRIED, H. E., HATFIELD, D. L. & HOWARD, M. T. 2015. Dietary Selenium Levels Affect Selenoprotein Expression and Support the Interferon-gamma and IL-6 Immune Response Pathways in Mice. *Nutrients,* 7**,** 6529-49.
TUJEBAJEVA, R. M., COPELAND, P. R., XU, X. M., CARLSON, B. A., HARNEY, J. W., DRISCOLL, D. M., HATFIELD, D. L. & BERRY, M. J. 2000. Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep,* 1**,** 158-63.
TURANOV, A. A., EVERLEY, R. A., HYBSIER, S., RENKO, K., SCHOMBURG, L., GYGI, S. P., HATFIELD, D. L. & GLADYSHEV, V. N. 2015. Regulation of Selenocysteine Content of Human Selenoprotein P by Dietary Selenium and Insertion of Cysteine in Place of Selenocysteine. *PLoS One,* 10**,** e0140353.
TURANOV, A. A., LOBANOV, A. V., FOMENKO, D. E., MORRISON, H. G., SOGIN, M. L., KLOBUTCHER, L. A., HATFIELD, D. L. & GLADYSHEV, V. N. 2009. Genetic code supports targeted insertion of two amino acids by one codon. *Science,* 323**,** 259-61.
TURANOV, A. A., LOBANOV, A. V., HATFIELD, D. L. & GLADYSHEV, V. N. 2013. UGA codon position-dependent incorporation of selenocysteine into mammalian selenoproteins. *Nucleic Acids Res,* 41**,** 6952-9.
TURANOV, A. A., MALINOUSKI, M. & GLADYSHEV, V. N. 2012. Selenium and Male Reproduction. *In:* HATFIELD, D. L., BERRY, M. J. & GLADYSHEV, V. N. (eds.) *Selenium: Its Molecular Biology and Role in Human Health.* New York, NY: Springer New York.

TURANOV, A. A., XU, X. M., CARLSON, B. A., YOO, M. H., GLADYSHEV, V. N. & HATFIELD, D. L. 2011. Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. *Adv Nutr,* 2**,** 122-8.

UMYSOVA, D., VITOVA, M., DOUSKOVA, I., BISOVA, K., HLAVOVA, M., CIZKOVA, M., MACHAT, J., DOUCHA, J. & ZACHLEDER, V. 2009. Bioaccumulation and toxicity of selenium compounds in the green alga Scenedesmus quadricauda. *BMC Plant Biol,* 9**,** 58.

VARGAS-RODRIGUEZ, O., ENGLERT, M., MERKURYEV, A., MUKAI, T. & SÖLL, D. 2018. Recoding of the selenocysteine UGA codon by cysteine in the presence of a non-canonical tRNACys and elongation factor SelB. *RNA Biology,* 15**,** 471-479.

VERES, Z., KIM, I. Y., SCHOLZ, T. D. & STADTMAN, T. C. 1994. Selenophosphate synthetase. Enzyme properties and catalytic reaction. *J Biol Chem,* 269**,** 10597-603.

VILARDELL, J., YU, S. J. & WARNER, J. R. 2000. Multiple functions of an evolutionarily conserved RNA binding domain. *Mol Cell,* 5**,** 761-6.

VOGEL, M., FISCHER, S., MAFFERT, A., HÜBNER, R., SCHEINOST, A. C., FRANZEN, C. & STEUDTNER, R. 2018. Biotransformation and detoxification of selenite by microbial biogenesis of selenium-sulfur nanoparticles. *Journal of Hazardous Materials,* 344**,** 749-757.

WALCZAK, R., WESTHOF, E., CARBON, P. & KROL, A. 1996. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *Rna,* 2**,** 367-79.

WANG, H., MCMANUS, J. & KINGSFORD, C. 2017. Accurate Recovery of Ribosome Positions Reveals Slow Translation of Wobble-Pairing Codons in Yeast. *J Comput Biol,* 24**,** 486-500.

WANG, K. T., WANG, J., LI, L. F. & SU, X. D. 2009. Crystal structures of catalytic intermediates of human selenophosphate synthetase 1. *J Mol Biol,* 390**,** 747-59.

WANG, W. X., MENG, J. & WENG, N. 2018. Trace metals in oysters: molecular and cellular mechanisms and ecotoxicological impacts. *Environ Sci Process Impacts,* 20**,** 892-912.

WANG, Y., CHEN, S., YAN, Z. & PEI, M. 2019. A prospect of cell immortalization combined with matrix microenvironmental optimization strategy for tissue engineering and regeneration. *Cell Biosci,* 9**,** 7.

WASCHULEWSKI, I. H. & SUNDE, R. A. 1988. Effect of Dietary Methionine on Utilization of Tissue Selenium from Dietary Selenomethionine for Glutathione Peroxidase in the Rat. *The Journal of Nutrition,* 118**,** 367-374.

WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M. A., CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* 25**,** 1189-1191.

WEINER, A. M. & WEBER, K. 1971. Natural read-through at the UGA termination signal of Q-beta coat protein cistron. *Nat New Biol,* 234**,** 206-9.

WEISS, R. B., DUNN, D. M., DAHLBERG, A. E., ATKINS, J. F. & GESTELAND, R. F. 1988. Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in Escherichia coli. *Embo j,* 7**,** 1503-7.

WEISS, R. B., HUANG, W. M. & DUNN, D. M. 1990. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell,* 62**,** 117-26.

WILLS, N. M., O'CONNOR, M., NELSON, C. C., RETTBERG, C. C., HUANG, W. M., GESTELAND, R. F. & ATKINS, J. F. 2008. Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *Embo j,* 27**,** 2533-44.

WOLFE, M. D., AHMED, F., LACOURCIERE, G. M., LAUHON, C. T., STADTMAN, T. C. & LARSON, T. J. 2004. Functional diversity of the rhodanese homology domain: the Escherichia coli ybbB gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. *J Biol Chem,* 279**,** 1801-9.

WOLFFRAM, S., BERGER, B., GRENACHER, B. & SCHARRER, E. 1989. Transport of Selenoamino Acids and their Sulfur Analogues across the Intestinal Brush Border Membrane of Pigs. *The Journal of Nutrition,* 119**,** 706-712.

WOLIN, S. L. & WALTER, P. 1988. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *Embo j,* 7**,** 3559-69.

WU, S., MARIOTTI, M., SANTESMASSES, D., HILL, K. E., BACLAOCOS, J., APARICIO-PRAT, E., LI, S., MACKRILL, J., WU, Y., HOWARD, M. T., CAPECCHI, M., GUIGO, R., BURK, R. F. & ATKINS, J. F. 2016. Human selenoprotein P and S variant mRNAs with different numbers of SECIS elements and inferences from mutant mice of the roles of multiple SECIS elements. *Open Biol,* 6.

XIA, Y., HILL, K. E., BYRNE, D. W., XU, J. & BURK, R. F. 2005. Effectiveness of selenium supplements in a low-selenium area of China. *Am J Clin Nutr,* 81**,** 829-34.

XIA, Y., HILL, K. E., LI, P., XU, J., ZHOU, D., MOTLEY, A. K., WANG, L., BYRNE, D. W. & BURK, R. F. 2010. Optimization of selenoprotein P and other plasma selenium biomarkers for the assessment of the selenium nutritional requirement: a placebo-controlled, double-blind study of selenomethionine supplementation in selenium-deficient Chinese subjects. *Am J Clin Nutr,* 92**,** 525-31.

XU, G. & JIANG, Y. Selenium and the prevalence of Keshan and Kaschin-Beck diseases in China. Proceedings of the 1st international symposium on geochemistry and health. Held at the Royal Society, London, 1985. 16-17.

XU, X. M., CARLSON, B. A., IRONS, R., MIX, H., ZHONG, N., GLADYSHEV, V. N. & HATFIELD, D. L. 2007a. Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochem J,* 404**,** 115-20.

XU, X. M., CARLSON, B. A., MIX, H., ZHANG, Y., SAIRA, K., GLASS, R. S., BERRY, M. J., GLADYSHEV, V. N. & HATFIELD, D. L. 2007b. Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol,* 5**,** e4.

XU, X. M., TURANOV, A. A., CARLSON, B. A., YOO, M. H., EVERLEY, R. A., NANDAKUMAR, R., SOROKINA, I., GYGI, S. P., GLADYSHEV, V. N. & HATFIELD, D. L. 2010. Targeted insertion of cysteine by decoding UGA codons with mammalian selenocysteine machinery. *Proc Natl Acad Sci U S A,* 107**,** 21430-4.

XU, X. M., ZHOU, X., CARLSON, B. A., KIM, L. K., HUH, T. L., LEE, B. J. & HATFIELD, D. L. 1999. The zebrafish genome contains two distinct selenocysteine tRNA[Ser]sec genes. *FEBS Lett,* 454**,** 16-20.

YAMAO, F., MUTO, A., KAWAUCHI, Y., IWAMI, M., IWAGAMI, S., AZUMI, Y. & OSAWA, S. 1985. UGA is read as tryptophan in Mycoplasma capricolum. *Proceedings of the National Academy of Sciences,* 82**,** 2306-2309.

YANG, G. Q., WANG, S. Z., ZHOU, R. H. & SUN, S. Z. 1983. Endemic selenium intoxication of humans in China. *Am J Clin Nutr,* 37**,** 872-81.

YIM, S. H., TOBE, R., TURANOV, A. A. & CARLSON, B. A. 2018. Radioactive 75Se Labeling and Detection of Selenoproteins. *In:* CHAVATTE, L. (ed.) *Selenoproteins: Methods and Protocols.* New York, NY: Springer New York.

YORDANOVA, M. M., WU, C., ANDREEV, D. E., SACHS, M. S. & ATKINS, J. F. 2015. A Nascent Peptide Signal Responsive to Endogenous Levels of Polyamines Acts to Stimulate Regulatory Frameshifting on Antizyme mRNA. *J Biol Chem,* 290**,** 17863-78.

YU, C. H., NOTEBORN, M. H., PLEIJ, C. W. & OLSTHOORN, R. C. 2011. Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. *Nucleic Acids Res,* 39**,** 8952-9.

YU, C. H., TEULADE-FICHOU, M. P. & OLSTHOORN, R. C. 2014. Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res,* 42**,** 1887-92.

ZAPATA, F., WILSON, N. G., HOWISON, M., ANDRADE, S. C., JORGER, K. M., SCHRODL, M., GOETZ, F. E., GIRIBET, G. & DUNN, C. W. 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc Biol Sci,* 281**,** 20141739.

ZAVACKI, A. M., MANSELL, J. B., CHUNG, M., KLIMOVITSKY, B., HARNEY, J. W. & BERRY, M. J. 2003. Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus. *Mol Cell,* 11**,** 773-81.

ZHANG, G., FANG, X., GUO, X., LI, L., LUO, R., XU, F., YANG, P., ZHANG, L., WANG, X., QI, H., XIONG, Z., QUE, H., XIE, Y., HOLLAND, P. W., PAPS, J., ZHU, Y., WU, F., CHEN, Y., WANG, J., PENG, C., MENG, J., YANG, L., LIU, J., WEN, B., ZHANG, N., HUANG, Z., ZHU, Q., FENG, Y., MOUNT, A., HEDGECOCK, D., XU, Z., LIU, Y., DOMAZET-LOSO, T., DU, Y., SUN, X., ZHANG, S., LIU, B., CHENG, P., JIANG, X., LI, J., FAN, D., WANG, W., FU, W., WANG, T., WANG, B., ZHANG, J., PENG, Z., LI, Y., LI, N., WANG, J., CHEN, M., HE, Y., TAN, F., SONG, X., ZHENG, Q., HUANG, R., YANG, H., DU, X., CHEN, L., YANG, M., GAFFNEY, P. M., WANG, S., LUO, L., SHE, Z., MING, Y., HUANG, W., ZHANG, S., HUANG, B., ZHANG, Y., QU, T., NI, P., MIAO, G., WANG, J., WANG, Q., STEINBERG, C. E., WANG, H., LI, N., QIAN, L., ZHANG, G., LI, Y., YANG, H., LIU, X., WANG, J., YIN, Y. & WANG, J. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature,* 490**,** 49-54.

ZHANG, Y., BARANOV, P. V., ATKINS, J. F. & GLADYSHEV, V. N. 2005a. Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J Biol Chem,* 280**,** 20740-51.

ZHANG, Y., FOMENKO, D. E. & GLADYSHEV, V. N. 2005b. The microbial selenoproteome of the Sargasso Sea. *Genome Biol,* 6**,** R37.

ZHANG, Y. & GLADYSHEV, V. N. 2008. Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. *PLoS genetics,* 4**,** e1000095-e1000095.

ZHOU, X., WANG, Z., CHEN, J., WANG, W., SONG, D., LI, S., YANG, H., XUE, S. & CHEN, C. 2014. Increased levels of IL-6, IL-1β, and TNF-α in Kashin–Beck disease and rats induced by T-2 toxin and selenium deficiency. *Rheumatology International,* 34**,** 995-1004.

ZINONI, F., BIRKMANN, A., LEINFELDER, W. & BÖCK, A. 1987. Cotranslational insertion of selenocysteine into formate dehydrogenase from Escherichia coli directed by a UGA codon. *Proceedings of the National Academy of Sciences of the United States of America,* 84**,** 3156-3160.